

BI

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
3 June 2004 (03.06.2004)

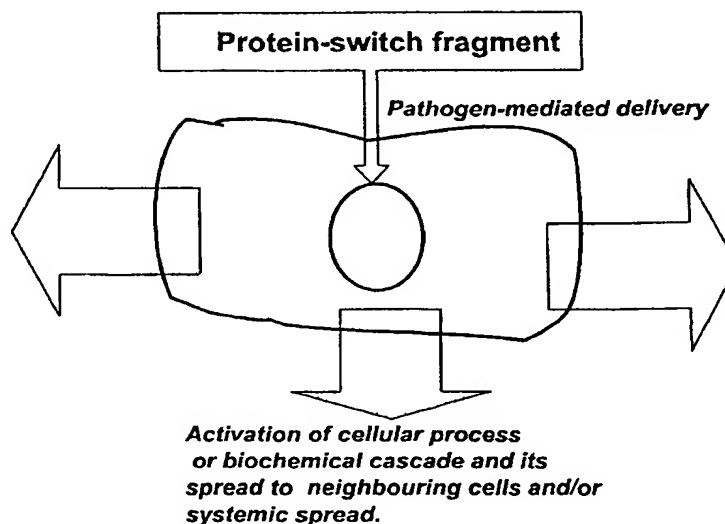
PCT

(10) International Publication Number
WO 2004/046359 A3

- (51) International Patent Classification⁷: C12N 15/82, A01H 5/00, C12N 1/21
- (21) International Application Number: PCT/EP2003/013016
- (22) International Filing Date: 20 November 2003 (20.11.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 102 54 167.1 20 November 2002 (20.11.2002) DE
- (71) Applicants (for all designated States except US): ICON GENETICS AG [DE/DE]; Maximilianstrasse 38/40, 80539 München (DE). ICON GENETICS, INC. [US/US]; 66 Witherspoon Street, Suite 134, Princeton, NJ 08542 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): WERNER, Stefan [DE/DE]; Rudolf-Haym-Str. 13, 06110 Halle (DE). MARILLONNET, Sylvestre [FR/DE]; Nietlebenerstrasse 20, 06126 Halle (DE). KLIMYUK, Victor [GB/DE]; Herderstrasse 10, 06114 Halle (DE). GLEBA, Yuri [UA/DE]; Mühlweg 37, 06114 Halle (DE).
- (74) Agent: WÄCHTERSCHÄUSER, Günter; Wächtershäuser & Hartz, Weinstrasse 8, 80333 München (DE).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published: — with international search report

[Continued on next page]

(54) Title: METHOD OF CONTROLLING GENE EXPRESSION IN PLANTS OR PLANT CELLS



(57) Abstract: A method of controlling a genetically-modified plant or plant cells, said method comprising the following steps: (a) providing a genetically-modified plant or plant cells, said plant or plant cells containing a heterologous nucleic acid encoding a first polypeptide containing or consisting of a first fragment of a protein, (b) introducing a second polypeptide into cells of said genetically-modified plant or plant cells, said second polypeptide containing (i) a second fragment of said protein and (ii) a peptide sequence enabling the introduction of said second polypeptide into cells of said genetically-modified plant or plant cells, whereby said first fragment and said second fragment jointly generate a predetermined function of said protein only when jointly present.

WO 2004/046359 A3

BEST AVAILABLE COPY



— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(88) Date of publication of the international search report:
26 August 2004

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 03/13016

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/82 A01H5/00 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/71701 A (NEW ENGLAND BIOLABS INC) 30 November 2000 (2000-11-30) page 2, line 14 - page 3, line 3; claim 8 page 5, line 15 - line 28 page 19, line 25 - page 27, line 11; figures 2-5 page 29 - page 45 page 90 - page 93	1,3, 10-18, 21,29,30
Y	----- -/--	2

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

24 June 2004

Date of mailing of the international search report

02/07/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

De Kok, A

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 03/13016

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/89283 A (HOOYKAAS PAUL JAN JACOB ; UNIVERSITEIT LEIDEN (NL); VERGUNST ANNETTE C) 29 November 2001 (2001-11-29) cited in the application page 1, line 24 - page 4, line 33	33
Y		2
A		3,16-18, 21,22
A	WO 00/52146 A (MAXYGEN INC (US);) 8 September 2000 (2000-09-08) page 5, line 12 - page 8, line 25 page 14, line 7 - page 26, line 3 page 41, line 1 - page 47, line 15	1,3,14, 15,22, 28,29, 31,32
A	WO 02/088369 A (ICON GENETICS AG) 7 November 2002 (2002-11-07) cited in the application page 4, last paragraph - page 11, paragraph 3; example 6	4-9, 23-25
A	US 2002/143142 A1 (LIN YAO-ZHONG ET AL) 3 October 2002 (2002-10-03) the whole document	16
A	WO 99/52563 A (UNIV CALIFORNIA (US)) 21 October 1999 (1999-10-21) cited in the application the whole document	19,20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 03/13016

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0071701	A	30-11-2000	AU 5039500 A	12-12-2000
			CA 2374497 A1	30-11-2000
			CN 1350582 T	22-05-2002
			EP 1183346 A1	06-03-2002
			JP 2003505012 T	12-02-2003
			WO 0071701 A1	30-11-2000
			US 2004096938 A1	20-05-2004
WO 0189283	A	29-11-2001	NL 1015252 C2	20-11-2001
			AU 6438901 A	03-12-2001
			CA 2409236 A1	29-11-2001
			CN 1476482 T	18-02-2004
			EP 1290200 A1	12-03-2003
			WO 0189283 A1	29-11-2001
			US 2004014025 A1	22-01-2004
WO 0052146	A	08-09-2000	AU 3391900 A	21-09-2000
			AU 3719500 A	21-09-2000
			CA 2362737 A1	08-09-2000
			CA 2364997 A1	08-09-2000
			EP 1165757 A1	02-01-2002
			EP 1165775 A2	02-01-2002
			WO 0052146 A2	08-09-2000
			WO 0052155 A2	08-09-2000
			US 6365377 B1	02-04-2002
			US 6531316 B1	11-03-2003
			US 6406910 B1	18-06-2002
			US 6358740 B1	19-03-2002
			US 6413745 B1	02-07-2002
WO 02088369	A	07-11-2002	DE 10121283 A1	31-10-2002
			CA 2439199 A1	07-11-2002
			WO 02088369 A1	07-11-2002
			EP 1385968 A1	04-02-2004
US 2002143142	A1	03-10-2002	US 6432680 B1	13-08-2002
			US 6248558 B1	19-06-2001
			AU 3462899 A	18-10-1999
			EP 1067949 A1	17-01-2001
			WO 9949879 A1	07-10-1999
WO 9952563	A	21-10-1999	AU 3644299 A	01-11-1999
			WO 9952563 A1	21-10-1999

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 December 2003 (11.12.2003)

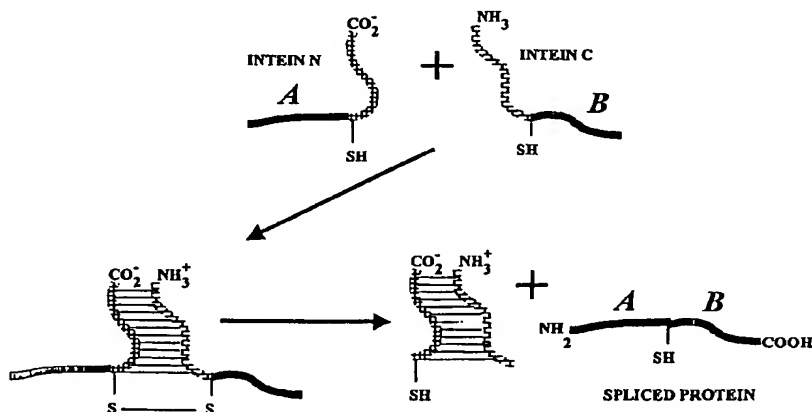
PCT

(10) International Publication Number
WO 03/102197 A1

- (51) International Patent Classification⁷: **C12N 15/82**
- (21) International Application Number: **PCT/EP03/02986**
- (22) International Filing Date: **21 March 2003 (21.03.2003)**
- (25) Filing Language: **English**
- (26) Publication Language: **English**
- (30) Priority Data:
102 24 214.3 31 May 2002 (31.05.2002) DE
102 24 980.6 5 June 2002 (05.06.2002) DE
- (71) Applicant (for all designated States except US): **ICON GENETICS AG [DE/DE]**; Maximilianstrasse 38/40, 80539 München (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **ICON GENETICS INC. [US/US]**; 66 Witherspoon Street, Suite 134, Princeton, NJ 08542 (US). **WERNER, Stefan [DE/DE]**; Rudolf-Haym-Str. 13, 06110 Halle (DE). **GIRITCH, Anatoly [UA/DE]**; Bergstr. 1, 06108 Halle (DE). **ELIBY, Serik [KZ/DE]**; Wittekindstr. 7, 06114 Halle (DE). **MARIL-LONNET, Sylvestre [FR/DE]**; Nietlebenerstr. 20, 06126
- Halle (DE). **KLIMYUK, Victor [GB/DE]**; Herderstr. 10, 06114 Halle (DE). **GLEBA, Yuri [UA/DE]**; Muehlweg 37, 06114 Halle (DE).
- (74) Agent: **HARTZ, Nikolai, F.**; Wächtershäuser & Hartz, European Patent and Trademark Attorneys, Weinstr. 8, 80333 München (DE).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report

[Continued on next page]

(54) Title: **TRANSGENIC PLANTS WITH CONTROLLED DISTRIBUTION OF A TRAIT TO PROGENY**



(57) Abstract: A process of producing a transgenic multi-cellular plants or parts thereof expressing a trait of interest, said trait having a controlled distribution of said trait to progeny, wherein said process comprises (i) producing a first plant or a cell thereof having in a first locus of a nuclear chromosome a first heterologous nucleotide sequence comprising a first fragment of a nucleotide sequence encoding said trait of interest, (ii) producing a second plant or a cell thereof having in a second locus of a nuclear chromosome homologous to said nuclear chromosome of step (i), a second heterologous nucleotide sequence comprising a second fragment of the nucleotide sequence encoding said trait of interest, and (iii) hybridising said first and said second plant or cells thereof to generate progeny exhibiting said functional trait of interest due to binding between a protein or polypeptide encoded by said first heterologous nucleotide sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence. Further, the invention provides a process of producing hybrid seeds for agriculture.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Transgenic plants with controlled distribution of a trait to progeny

FIELD OF THE INVENTION

The present invention relates to a process of producing hybrid seeds. The invention also relates to a process of producing a transgenic multicellular plant organism expressing a trait of interest and having a controlled distribution of said trait to progeny or to other organisms. The invention also relates to a process of producing a transgenic multicellular plant organism expressing two traits of interest, whereby said traits have a controlled distribution to progeny. Preferably, one of said traits is male sterility. Moreover, the invention relates to a process of producing hybrid seeds, notably for agricultural purposes. The invention further relates to a plant expressing a trait, whereby the distribution of said trait to progeny is controlled, i.e. the probability of transferring said trait to illicit progeny, notably by cross-pollination, is very low.

BACKGROUND OF THE INVENTION

The commercial use of genetically engineered crop species has caused concerns about the possible transfer of transgenes and traits encoded by transgenes from genetically modified plants (GM plants) into landraces, wild relatives or other non-GM plant varieties or related crop species (Ellstrand, N. C., 2001, *Plant Physiol.* 125, 1543-1545; Quist & Chapela, 2001, *Nature*, 414, 541-543), which could change the ecological balance in the affected ecosystems or lead to other, first of all, socioeconomic problems. Additionally, there is a certain fear that transgenes, especially antibiotic resistance genes used as transformation markers, can escape, through so-called horizontal transfer, into surrounding microorganisms (Chiter *et al.*, 2000, *FEBS Lett.*, 481, 164-168), thus modifying the microflora in an undesirable way.

Although many of these worries are not well justified scientifically (Christou, P., 2002, *Transgenic Res.*, 11, iii-v), the creation of safe and controlled transgene management systems is highly desirable, as it might prevent potential problems in the future and shall help to protect the germplasm of existing plant species in the most efficient way. In addition, there are problems caused by contamination of organically grown crops or non-GM crops with transgenic cultivars. This has a serious impact on the marketing of transgenic as well as non-transgenic crops, an issue which cannot be ignored by producers.

Unlike other products generated by humans, products created by biotechnology are potentially self-replicating machines. Therefore, any transgenic material created by current technology and released into the environment has a potential of persisting there for a very long time. Common practice of plant genetic engineering is based on the use of expression cassettes and vectors that contain continuous coding sequence for the gene of interest. Such expression cassettes are integrated into a host chromosome and upon hybridization or another genetic information exchange between a GM plant and another organism, whether licit or illicit, the expression cassette is transmitted with a high probability to the progeny or another recipient as a functional transcriptional unit.

WO00/52146 describes general ideas for encrypting a trait of interest by splitting gene(s) in two or more fragments and rejoining the fragments by trans-splicing after mating parental organisms, whereby the parental organisms provide said fragments. WO00/52146 does not go beyond general ideas. It does not contain an enabling disclosure on how these ideas can be reduced to practice. Notably, it does not contain an example. WO00/71701 describes assembly of a functional protein by intein-mediated protein trans-splicing/interaction for improving containment of a transgene encoding said protein. WO00/71701 does not describe bringing together fragments of a protein by mating parent organisms. Further, the frequency of transmission of transgene according to WO00/71701 is not sufficiently low for large scale applications like agriculture, notably when a transgene provides a selective advantage.

WO0116287 relates to the creation of allelic position for transgenes, whose expression determines a phenotype, with the aim that the transgenes segregate to different gametes. This patent application does not address the problem of controlling movement of transgenes, but rather trait generation, specifically male-sterility, encoded by at least two transgenes. Further, it does not mention intein-mediated trans-splicing. Moreover, this application does not describe control over trait movement by splitting a trait-encoding gene in two or more fragments.

Trait assembly from parts encoding the trait is not of high value without knowing how to achieve the most favorable positions of the encoding fragments in practically the most feasible way, in order to provide the strictest control over undesired transmission of said trait. For large scale applications like for agriculture, biological safety requires that undesired transmission of a transgene is reduced to a frequency of practically zero.

Crop plants expressing as a trait of interest male or female sterility are widely used for hybrid seed production. Hybrid crops have on average 20% yield advantage over inbred varieties and production of hybrid seeds is a large industry. Many different technologies are used to produce hybrid seeds (for review see: Perez-Prat E. & van Lookeren Campagne, MM, 2002, *Trends Plant Sci.*, 7, 199-202). These technologies can be conditionally divided into at least four groups according to the pollination control mechanism: mechanical, chemical, genetic and transgenic. However, one critical requirement is common for all these technologies: ideally, a 100% male sterile line should be used for the hybridization process and 100% male fertility restoration in F_1 progeny should be achieved. Such stringent requirements are absolutely necessary for producing hybrid seeds free of contamination with selfed seeds.

The current methods of hybrid seed production are unsatisfactory in the above respect. These processes are either expensive, as in the case of mechanical de-tasselling (castration) of corn, or "leaky" as in the case of genetic approaches or both as in the case of chemical treatment-based method (e.g. US4569688).

Genetic approaches preferably include the use of lines with cytoplasmic male sterility (CMS) mutants and fertility restorers (e.g. WO02098209). Transgenic approaches use predominantly plants with genetically engineered nuclear male sterility (NMS) or CMS and fertility restoration in F_1 progeny (WO8910396; US5530191; US6255564; WO9832325; WO9201799; US6392119; WO0116287). These approaches also require the use of a so-called maintainer line in order to propagate and maintain the male-sterile line.

The transgenic systems built on one transgene providing for male sterility and another transgene carrying the function of restoring male fertility (e.g. US62555640) guarantee neither complete restoration of male fertility in hybrid progeny nor complete elimination of potentially negative effects of the transgene providing for male sterility on the general health of said progeny. In other words these systems are leaky. In addition, none of the systems mentioned above offers a convenient way of producing and maintaining the male-sterile line. This is an important element of any genetically engineered system for hybrid seed production, as the successful application of such a system for large-scale production depends on whether the male-sterile female parent line can be propagated in an economical and efficient way. In other words, currently there is no universal, reliable and economical system for hybrid seed production, which integrates all requirements necessary for maintenance of the original lines, hybridization process, restoration of male fertility in hybrid progeny and at the same time has high biological safety parameters, e.g. provides for tight control over transgene segregation. A

general scheme of hybrid seeds production using currently existing genetic/transgenic approaches is shown in Figure 12.

In the present invention, we describe a new process of producing hybrid seeds (Fig. 13) which has all necessary characteristics to match the requirements of an ideal hybridization system. A comparison of the hybrid seed production system of the invention with prior art methods is presented in Table 1.

It is therefore an object of the invention to provide a process of producing a transgenic plant expressing a trait of interest, notably male sterility, whereby distribution of said trait to progeny is strictly controlled and occurs with low probability.

It is a further object of the invention to provide a process of producing a biologically safe transgenic plant, notably a male sterile plant, that expresses a trait of interest, whereby gene fragments encoding said trait are positioned such that undesired transmission of said trait occurs with low probability.

It is a further object of the invention to provide a process of positioning transgenic DNA sequences on homologous chromosomes, notably in the same locus of homologous chromosomes of a multi-cellular organism.

It is also an object of the invention to provide a process of producing a male sterile plant line.

It is another object of the invention to provide a universal and environmentally safe process of producing hybrid seeds using a sterile plant line, whereby complete fertility restoration occurs in said hybrid seeds.

GENERAL DESCRIPTION OF THE INVENTION

The invention provides a process of producing a transgenic multi-cellular plant organism or parts thereof expressing a trait of interest and having a controlled distribution of said trait to progeny, wherein said process comprises

- (i) producing a first plant or a cell thereof having in a first locus of a nuclear chromosome a first heterologous nucleotide sequence comprising a first fragment of a nucleotide sequence encoding said trait of interest,
- (ii) producing a second plant or a cell thereof having in a second locus of a nuclear chromosome homologous to said nuclear chromosome of step (i), a second heterologous nucleotide sequence comprising a second fragment of the nucleotide sequence encoding said trait of interest, and
- (iii) hybridising said first and said second plant or cells thereof to generate progeny exhibiting said functional trait of interest due to binding between a protein or polypeptide encoded by said first heterologous nucleotide sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence. Said binding preferably involve protein trans-splicing.

Said multi-cellular plant organisms or said parts produced by the above process may express two traits of interest, a trait (1) and a trait (2), both traits having a controlled distribution to progeny.

The inventors of this invention have developed for the first time a method of rendering transgenic plants environmentally safe in that the transgene or a trait of interest expressed by said plant has a controlled distribution to progeny of said plant. The invention solves a major problem of biotechnology, notably of plant biotechnology, since transfer of a transgene from a GM plant to other organisms can now be effectively controlled and limited. Transfer of a transgene to other organisms includes transfer to sexual progeny by cross-pollination as well as lateral gene transfer. The above processes make obtainable genetically modified multi-cellular plants with a controlled containment of a trait of interest.

In an important embodiment, said trait of interest is male or female sterility, preferably male sterility. In this case, the transgenic multi-cellular plant organism of the invention may be used for hybrid seed production by crossing with another plant that is male fertile or female fertile, respectively. The hybrid seeds produced using the transgenic multi-cellular plant of the invention may be 100 % fertile due to a controlled distribution of the sterility trait to progeny. In

a particularly preferred embodiment, said transgenic multi-cellular plant of the invention may express two traits of interest, a male sterility trait and a herbicide resistance trait, what makes amenable a novel process of producing hybrid seeds with several advantages over prior art processes (see below).

In the process of the invention, the nucleotide sequence encoding (or involved in) said trait is split into two or more fragments. Preferably, said nucleotide sequence is split into two fragments of said nucleotide sequence, thus obtaining a 5' and a 3' part of the nucleotide sequence. Said 5' part corresponds essentially to said first fragment. Said 3' part corresponds essentially to said second fragment. Said nucleotide sequence is typically a coding sequence (or an open reading frame) of a protein involved in said trait. However, said nucleotide sequence may contain one or more introns. To obtain said fragments, said nucleotide sequence is preferably split such that each obtained fragment, upon expression, is incapable of generating said trait in the absence of the other fragment. Each fragment contains a sequence portion necessary for the function of the protein involved in said trait. For example, if said protein involved in said trait is an enzyme, each fragment preferably contains amino acids necessary for catalysis or substrate binding of the enzyme. A protein involved or encoding a trait may be split into said fragments in many different ways provided that expression of said trait requires all said fragments and binding thereof to each other. Structural and functional information known about the protein involved in said trait may be helpful for finding a suitable splitting site of said nucleotide sequence. In any case, one can easily test experimentally whether a fragment generated by splitting a nucleotide sequence at a randomly chosen site is capable of expressing a trait encoded by said nucleotide sequence. The following description focuses on the preferred embodiment, wherein said nucleotide sequence encoding said trait is split into two fragments.

Expression of said trait requires the presence of both said fragments in the same plant, preferably in the same cells thereof. Expression of said trait further requires transcription and translation of said first and said second fragment and binding of the translation products of said fragments to each other with or without peptide bond formation. Preferably, said binding involves peptide bond formation between said fragments.

The first fragment is incorporated into a first heterologous nucleotide sequence, the second fragment is incorporated into a second heterologous nucleotide sequence. Preferably, said heterologous nucleotide sequences are DNA sequences.

Preferably, said first and said second heterologous nucleotide sequence further codes for a first and a second binding polypeptide, respectively, that renders said polypeptides encoded by said first and said second heterologous nucleotide sequences capable of said binding. Each binding polypeptides is preferably expressed as a protein fusion with the polypeptide encoded by said first or said second fragment.

Said polypeptide or protein encoded by said first heterologous nucleotide sequence comprises, preferably consists of, a first binding polypeptide and a polypeptide encoded by said first fragment. Said polypeptide or protein encoded by said second heterologous nucleotide sequence comprises, preferably consists of, a second binding polypeptide and a polypeptide encoded by said second fragment.

After transcription and translation, each of said polypeptides or proteins has at least the following two functions:

- (i) providing a part of the protein involved in said trait;
- (ii) the capability of binding to the polypeptide or protein encoded by the other fragment.

Amino acid sequence portions responsible for said functions (i) and (ii) may or may not overlap.

Said binding may or may not involve peptide bond formation between said proteins or polypeptides encoded by said first and second heterologous nucleotide sequences. Without peptide bond formation, said binding polypeptides may bind to each other by affinity. In this case, said binding polypeptides may be polypeptides known to bind to each other e.g. from naturally occurring binding domains of protein complexes. Preferably, said binding polypeptides involved in said binding affinity or at least one of them can be artificially engineered. Said binding polypeptides may e.g. be the components of an antigen-antibody pair. Further, said binding polypeptides may be selected artificially using e.g. random peptides phage display libraries (for review see: Barbas CF., 1993, *Curr Opin. Biotechnol.*, 4, 526-530; Irving *et al.*, 2001, *Curent Opin. Chem. Biol.*, 5:314-324; Hoogenboom HR, 1997, *Trends Biotechnol.*, 15: 62-70) or yeast two-hybrid system (for review see Fields & Sternglanc, 1994, *Trends Genet.*, 10, 286-292; Bartel & Fields., 1995, *Methods Enzymol.*, 254: 241-263). Further, they may be intein fragments that may have been rendered non-functional for intein splicing.

In an important embodiment, said binding comprises peptide bond formation between said protein or polypeptides encoded by said first and second heterologous nucleotide sequences. Peptide bond formation between the polypeptides encoded by said fragments is preferred. Said binding is or comprises preferentially intein-mediated trans-splicing. For this purpose, said first and said second heterologous nucleotide sequences further code for proteins or polypeptides capable of protein trans-splicing. By said trans-splicing, the proteins and polypeptides encoded by said first and said second fragments may be linked by peptide bond formation. In this embodiment, said binding polypeptides are preferably derived from an intein capable of trans-splicing. Trans-splicing inteins may be selected from the nucleolar and organellar genomes of different organisms including eukaryotes, archaebacteria and eubacteria. Inteins that may be used for performing this invention are listed at <http://www.neb.com/neb/inteins.html>. Also, an intein mentioned in a reference cited herein may be used. The choice of the intein might depend on the consensus sequences as well as the conditions required for efficient trans-splicing.

For engineering said heterologous nucleotide sequences, the nucleotide sequence coding for an intein may be split into a 5' and a 3' part that code for the 5' and the 3' intein (as denoted herein), respectively. Sequence portions not necessary for intein splicing (e.g. a homing endonuclease domain) may be deleted. The intein coding sequence is split such that the 5' and the 3' inteins are capable of trans-splicing. Regarding a suitable splitting site of the intein coding sequence, the considerations published by Southworth et al. (EMBO J. (1998) 17, 918-926) may be followed. The capability of the 5' and the 3' inteins for trans-splicing may of course be tested experimentally, e.g. as described by Southworth et al. (ibid). Experimental testing may be done by trans-splicing. Experimental testing of intein portions that can be deleted without compromising trans-splicing functionality may be done by trans-splicing or by cis-splicing.

The 5' intein corresponds essentially to the first binding polypeptide. The 3' intein corresponds essentially to the second binding polypeptide. For engineering said heterologous nucleotide sequences, the 5' intein coding sequence is linked to the 3' end of said first fragment. The 3' intein coding sequence is linked to the 5' end of said second fragment. Notably in the vicinity of the linking site, nucleotides and/or codons (amino acids) may be changed to achieve a desired trans-splicing functionality.

Said first heterologous nucleotide sequence thus may comprise: said first fragment, said first binding polypeptide, regulatory sequences for transcription (e.g. promoter, 3' transcription termination sequence) and for translation. Said second heterologous nucleotide sequence may comprise: said second fragment, said second binding polypeptide, regulatory sequences for transcription (e.g. promoter, 3' transcription termination sequence) and for translation. Further, it may contain a selectable and/or a counter-selectable marker needed for producing said first and/or said second plant and sequences recognised by a site-specific recombinase or transposon sequences (cf. below).

The process of the invention may also be used to assemble two or more traits, notably by trans-splicing. However, different intein systems should be used for the assembly of each trait in order to avoid trait mis-splicing due to the universal nature of interaction between intein parts, which is independent of attached protein fragment destined for trans-splicing.

In the process of the invention, said first plant or cells thereof may be produced by introducing said first heterologous nucleotide sequence into a precursor plant or cells thereof. Said second plant or cells thereof may be produced by introducing said second heterologous nucleotide sequence into a precursor plant or cells thereof. Said introducing may be done according to methods generally known in the art. Preferably, both heterologous nucleotide sequences are stably incorporated into a chromosome of the nuclear genome of the first and the second plant. Said first and said second plants obtained thereby are preferably made homozygous with respect to the respective heterologous nucleotide sequences according to procedures known in the art, notably by selfing. Said first and said second plants belong preferably to the same family, more preferably to the same genus, and most preferably to the same species of organisms.

The invention provides multi-cellular plants (and parts thereof like seeds) expressing a trait of interest and having a controlled distribution of said trait to progeny, whereby a protein involved in said trait is generated by binding, notably by trans-splicing, polypeptides encoded by said heterologous sequences. Said polypeptides are encoded on homologous chromosomes of said organism in a first and a second heterologous nucleotide sequence.

In principle, several relative locations of said first and said second heterologous nucleotide sequences and the respective fragments exist in the transgenic plant of the

invention. Said first and said second heterologous sequences in said transgenic plant of the invention should be positioned such that they segregate as unlinked loci. Said unlinked loci are preferably positioned so as to minimize meiotic recombination or crossing-over and creation of linkage between said loci.

Possible relative locations of said first and said second heterologous nucleotide sequences and said fragments contained therein are generally shown in Fig. 2B using a diploid organism as an example.

In case I of Fig. 2B, said first and said second fragments are located on the same chromosome, i.e. they are physically linked on the same DNA molecule but are separated from each other by chromosome sequences native to the organism. The fragments will belong to different transcriptional units. Since crossing-over in meiosis may lead to separation of the fragments (or the heterologous sequences containing the fragments), the probability of transferring the trait encoded by both fragments to progeny is reduced compared to the conventional case, where the trait is encoded by a continuous coding sequence.

In case II (see Fig. 2B), said first and said second fragment are located on different heterologous chromosomes. The frequency of inheriting said trait encoded by the two fragments on different chromosomes upon self-crossing is about 50% and upon crossing with an organism not carrying any of these fragments 25%. In prior art cases I and II, the probability of transferring both fragments to progeny or to other organisms is too high for practical purposes, notably if the trait encoded in said fragments provides an advantage for survival or propagation. These cases do not represent biologically safe cases of a transgenic plant.

The inventors of this invention have found that the frequency of transferring said trait to progeny (upon crossing with plants not having said trait) and to other organisms can be enormously reduced when said fragments are located on homologous chromosomes as schematically shown in Fig. 2B, case III and IV.

In case III (Fig. 2B), the two fragments are present at different loci on homologous chromosomes, i.e. are linked in repulsion. The closer the fragments are located, the lower the frequency of recombination between said loci and, consequently, transferring the trait to progeny as the result of cross-hybridisation. In the most preferred case (case IV in Fig. 2B), the fragments are located in the same locus on homologous chromosomes. Thus, the trait reliably segregates in cross-progeny (hybrid progeny) of the multi-cellular plant of the invention.

Such relative locations of said first and said second heterologous nucleotide sequences on homologous chromosomes of the plant of the invention are achieved by hybridising, notably crossing, said first and said second plant or cells thereof. Said first and said second plant may be obtained by methods known in the art. Further possibilities are disclosed in the following.

In one embodiment, many transformants are produced with said first as well as with said second heterologous nucleotide sequence. Then, the chromosome having said heterologous sequence incorporated as well as the location of the transformed sequence in the chromosome may be determined by genetic or molecular biological methods. Next, a transformed plant or cell clone thereof having said first heterologous nucleotide sequence at a suitable location may be selected. Then, a transformed plant or cell clone thereof having said second heterologous nucleotide sequence at a suitable location relative to said first sequence may be selected. Thereby, a suitable pair of first and second plants may be chosen.

In a second embodiment, targeted integration into a desired locus of a desired chromosome is employed making use of homologous recombination. Preferably, targeted integration is done using a multi-cellular plant having a targeting site pre-integrated into a chromosome in combination with site-specific recombination. The latter approach is particularly useful for introducing said first and said second heterologous nucleotide sequence into the same locus of the same chromosome, as the same starting organism line having a pre-integrated targeting site may be used for transforming said first and said second heterologous nucleotide sequences. Targeted integration is described e.g. in international patent application PCT/EP02/03266 (WO02/077246). Methods of creating sites for targeted integration in plants with different expression profiles are described in PCT/US02/11924. Methods of improving the efficiency of site-targeted integration is described e. g. in international patent application PCT/EP02/03266.

Alternatively, said first heterologous nucleotide sequence can be incorporated into a chromosome of the nuclear genome of the first organism and said second heterologous nucleotide sequence can be incorporated into the plastid or mitochondrial genome of the same or another organism. However, incorporation of both heterologous nucleotide sequences into nuclear chromosomes is preferred.

Preferred methods of producing said first and said second plant are schematically depicted on the right hand side („Excision“) of Fig. 2E and in Figs. 5 to 8. In these preferred methods, steps (i) and (ii) of claim 1 are carried out by

- (a) introducing a parent heterologous nucleotide sequence comprising said first and said second heterologous nucleotide sequences into a nuclear chromosome of parent organisms or cells thereof,
- (b) optionally selecting organisms or cells thereof having said parent heterologous nucleotide sequence integrated in a desired chromosome or chromosome locus,
- (c) subsequently splitting said parent heterologous nucleotide sequence so that said first and said second heterologous nucleotide sequences are located on homologous chromosomes in different plant organisms or cells.

Said parent heterologous nucleotide sequence comprises said first and said second heterologous nucleotide sequence. Preferably, it further comprises sequences for excising said first and/or said second heterologous sequence (for details see below). Said introducing (a) may be done by any known transformation method (see below). *Agrobacterium*-mediated transformation preferred. Plants or cells carrying said parent heterologous nucleotide sequence may be selected using a selectable marker contained therein. Whole plants may be regenerated from transformed cells or tissue. Preferably, plants homozygous for said parent sequence are created.

A plant (or a group of plants) carrying said parent sequence may then be used for excising said first heterologous nucleotide sequence out of said parent sequence. Thus, said second plant may be obtained. Another plant (or group of plants) may be used for excising said second heterologous nucleotide sequence for obtaining said first plant. The heterologous sequences which are not excised are located in said first and said second plant in homologous chromosomes, notably in the same locus of said homologous chromosomes, i.e. in iso-loci.

The first and second plants or cells thereof thus obtained (or progeny thereof) are advantageously analysed for any reintegration of an excised heterologous nucleotide sequence into the genome e.g. by genetic or molecular biological techniques (e.g. by PCR and use of nucleotide probes for Southern hybridisation). Plants or cells thereof may then be selected that contain said heterologous nucleotide sequence reintegrated at a desired locus on a chromosome homologous to the chromosome harboring the heterologous nucleotide sequence that has not been excised. Thus, the transgenic plant of the invention may directly be obtained. Preferably, plants or cells thereof that are free of the excised heterologous nucleotide sequence

are selected. Said selection may comprise analysis by genetic or molecular biological techniques. Preferably, said selection is supported by a counter-selectable marker in the heterologous sequence to be excised. Said first and said second plant are preferably made homozygous for said heterologous sequence that has not been excised.

Said excising may e.g. be done using site-specific recombinases (cf. Fig. 2E). It is highly convenient that said excising is done using transposons, notably non-autonomous transposons (i.e. a transposon not encoding the respective transposase). For the latter embodiment, said first and/or said second heterologous nucleotide sequence in said parent heterologous nucleotide sequence is/are embedded in such a transposon. Said excision comprises providing a transposase for said transposon. Notably,

- (A) said first heterologous nucleotide sequence in said parent heterologous nucleotide sequence is contained in a first transposon and said second heterologous nucleotide sequence is contained in a second transposon and
- (B) said first heterologous nucleotide sequence is excised by providing a first transposase functional with said first transposon and said second heterologous nucleotide sequence is excised by providing a second transposase functional with said second transposon.

Said first and said second transposons in said parent heterologous nucleotide sequence preferably overlap such that excision of said first or said second heterologous nucleotide sequence leads to disruption of said second or said first non-autonomous transposon, respectively. Overlapping transposons may conveniently be used with a selectable and a counter-selectable marker in the overlapping region as depicted in Fig. 7 and 8.

Further, said splitting of step (c) does not necessarily require different recombinases for said excising said first or said second heterologous nucleotide sequence. In a very convenient embodiment, said first heterologous nucleotide sequence in said parent heterologous nucleotide sequence is flanked by differing recombination sites of a site-specific integrase and said second heterologous nucleotide sequence in said parent heterologous nucleotide sequence is flanked by differing recombination sites of the same site-specific integrase (cf. Fig. 21 and 22), and step (c) is carried out by

- providing said site-specific integrase to said parent organism or cells thereof,
- selecting progeny of said parent organism or cells thereof containing said first heterologous nucleotide sequence but not said second heterologous nucleotide sequence, and

selecting progeny of said parent organism or cells thereof containing said second heterologous nucleotide sequence but not said first heterologous nucleotide sequence.

In step (iii) the process of the invention, said first and said second plants or cells thereof are then hybridised for obtaining the transgenic multi-cellular plant of the invention. Hybridising may be sexual crossing or fusion of cells of said plants. Cell fusion may be fusion of germ cells or of somatic cells. Preferably, hybridising involves pollination of plants or somatic cell fusion of protoplasts. Sexual crossing of plants is most preferred. Said hybridising brings said fragments encoding or being involved in said trait together in one plant or cells thereof such that said plant exhibits said trait of interest due to protein trans-splicing. Exhibiting said trait due to protein binding or protein trans-splicing means that binding or trans-splicing is a necessary condition for the expression of said trait of interest. The production of the transgenic organism of the invention may comprise further steps in addition to said hybridising. In the case of plants, examples of such further steps include: growing and harvesting seeds, seeding, and growing the plant of the invention. In the case of protoplast fusion, such further steps include: propagating the fused protoplasts to obtain colonies, regeneration of plants.

Controlled distribution of said trait to progeny means that the probability of transferring said trait to progeny is significantly reduced compared to conventional transgenic organisms that have a transgene involved in said trait of interest encoded in one locus of a chromosome, notably as a single transcriptional unit, or on heterologous chromosomes. The frequency of appearance of said trait in progeny upon crossing said transgenic multi-cellular plant of the invention with a plant devoid of said first and said second heterologous sequences is less than 10%, preferably less than 1 %, more preferably less than 0.1%, even more preferably less than 0.01%, and most preferably less than 0.001%. For comparison, the frequency of appearance of a transgene in progeny upon crossing a conventional transgenic (diploid) organism having said transgene in a single transcriptional unit and being heterozygous with respect to the transgene with another organism of the same species not having said transgene is about 50%. Whether a transgenic plant expressing a trait of interest fulfills the criteria of the invention regarding said frequency can be easily checked experimentally.

Herein, peptide bond means the amide linkage between the carboxyl group of one polypeptide and the amino group of another polypeptide. The linkage does not allow free

rotation and can occur in *cis* or *trans* configuration, the latter the most common in natural peptides, except for links to the amino group of proline, which are always *cis* (source: <http://www.mblab.gla.ac.uk/dictionary/>). Peptide bond formation can be achieved through intein-mediated trans-splicing.

In the process of the invention, transgenic multi-cellular plant organisms are produced. Among plants, crop plants including barley, oat, rye, wheat, *zea mays*, rice, millet, potato, oilseed rape, canola, tomato, cotton, sorghum, and tobacco are preferred. The processes of the invention may be applied to diploid and to polyploid plants.

Examples for traits expressible according to the invention, notably in plants, are male sterility, herbicide resistance, insecticide resistance, selectable marker, a counter-selectable marker, organism morphology, seed content, seed stability, climate adaption, vitamine content, carbohydrate content and composition, fat content and composition etc. Further, said trait may be expression of a protein of interest, notably a pharmaceutical protein. Examples for such proteins are given below. In one case (cf. example 1), reporter gene is expressed in a plant of the invention. In another example of this invention (example 2) EPSPS (5-enolpyruvylshikimate-3-phosphate synthase) gene conferring herbicide resistance, e.g. glyphosate tolerance, is expressed. Said multi-cellular plants and said transgenic multi-cellular plants of the invention may be further genetically or transiently modified e.g. for providing functions necessary for said trans-splicing and/or said expressing of the trait of interest. Further, a second transgene involved in expression of said trait of interest or of a different trait may be expressed.

The process of the invention may be used for a wide variety of applications. It may e.g. be used for expressing a trait of interest in said transgenic organism. Said trait may be any property of said organism, whether encoded by a single or by several genes. Said trait may be caused by expression of at least one protein. Two or more proteins may be necessary for said trait. In this case, it may be sufficient to control the expression of only one protein as described herein. It is, however, environmentally safer to control all the proteins producing a trait by the processes of the invention.

A highly important application of said process is the production of hybrid seeds for generating plants for agricultural purposes or for protein production in said plants, whereby said plants have a controlled distribution of a trait to progeny. Said hybrid seeds allow the generation of plants expressing a trait of interest that is neither expressed in a parental line and quickly

segregates in progeny.

Producing plants or cells thereof expressing two traits of interest with controlled distribution of said traits to progeny

The transgenic multi-cellular plants or parts thereof produced according to the invention may be made to express two (or more) traits of interest, whereby both traits may have a controlled distribution to progeny as defined above. For the preferred case of two such traits of interest, these are referred to in following as trait (1) and trait (2). The above description regarding said trait of interest may apply to said trait (1) or to said trait (2). Preferably, it applies to said trait (1) and to said trait (2). However, expression of trait (1) or trait (2) may depend on RNA trans-splicing of mRNA expression products of said first and said second heterologous nucleotide sequence. Translation of the trans-spliced RNA may in this case generate one of said traits (1) or (2). RNA trans-splicing is described in detail in WO02/96192 and in references cited therein. It is also possible to expressed two or more traits via RNA trans-splicing.

Preferably, the progeny generated in step (iii) of the process of the invention (i.e. the transgenic multi-cellular plants or parts thereof according to the invention) exhibits trait (1) and trait (2) due to binding between a protein or polypeptide encoded by said first heterologous nucleotide sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence. Further, said progeny may exhibit trait (1) or trait (2) due to intein-mediated trans-splicing. Further, said progeny may exhibit trait (1) and trait (2) due to intein-mediated trans-splicing.

In the process of producing a multi-cellular plant or parts thereof expressing two traits of interest, steps (i) and (ii) may be carried out similarly as described above in detail for one trait. The plant produced in step (i) (plant A1 in Fig. 13) may contain a (first) fragment of a nucleotide sequence encoding trait (1) and a (first) fragment of a nucleotide sequence encoding trait (2). The plant produced in step (ii) (plant A2 in Fig. 13) may contain another (a second) fragment of a nucleotide sequence encoding trait (1) and another (a second) fragment of a nucleotide sequence encoding trait (2). Said first fragments (of trait (1) and of trait (2)) in the plant produced in step (i) may be on the same or on different chromosomes. Similarly, said second fragments (of trait (1) and of trait (2)) in the plant produced in step (ii) may be on the same or on different chromosomes. It is preferred that said first fragments are on the same chromosome and that said second fragments are on the same chromosomes. More preferably, said first fragments are in the same locus of a chromosome and said second fragments are in the same

locus of a chromosome. Most preferably, the locus having said first fragments of said first plant and said locus having said second fragments of said second plant are the same loci on homologous chromosomes, i.e. are iso-loci.

In a preferred embodiment, said first fragment of a nucleotide sequence encoding trait (1) and said first fragment of a nucleotide sequence encoding trait (2) are contained in said first heterologous nucleotide sequence of the invention. Similarly, said second fragment of a nucleotide sequence encoding trait (1) and said second fragment of a nucleotide sequence encoding trait (2) are contained in said second heterologous nucleotide sequence of the invention. Step (iii) may then comprise hybridising said first and said second plant or cells thereof to generate progeny exhibiting trait (1) and trait (2), whereby exhibiting of trait (1) is due to binding between a protein or polypeptide encoded by said first heterologous nucleotide sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence.

In the aforementioned preferred embodiment, a strictly controlled distribution of trait (1) and of trait (2) in the plant produced by the process of the invention can conveniently be achieved, if said first and said second heterologous nucleotide sequence are located in iso-loci in said first and said second plant. Therefore, progeny obtained by crossing said transgenic multi-cellular plant of the invention that expresses said two traits of interest with another plant not containing said fragments will express neither trait (1) nor trait (2).

Steps (i) and (ii) are preferably carried out by

- (a) introducing a parent heterologous nucleotide sequence comprising said first and said second heterologous nucleotide sequences into a nuclear chromosome of parent organisms or cells thereof,
- (b) optionally selecting organisms or cells thereof having said parent heterologous nucleotide sequence integrated in a desired chromosome or chromosome locus,
- (c) subsequently splitting said parent heterologous nucleotide sequence so that said first and said second heterologous nucleotide sequences are located on homologous chromosomes in different plant organisms or cells,

whereby said first heterologous nucleotide sequence of said parent nucleotide sequence contains the first fragment of trait (1) and the first fragment of trait (2), and said second heterologous nucleotide sequence of said parent nucleotide sequence contains the second fragment of trait (1) and the second fragment of trait (2). Said splitting of step (c) may be carried out as described above, whereby plant A1 and plant A2 may be obtained. Preferably, said

plants produced in step (i) and in step (ii) are selfed for rendering them homozygous for said first and/or said second heterologous nucleotide sequence.

Examples for trait (1) and for trait (2) may be those given above.

Process of hybrid seed production

In an embodiment of utmost importance, trait (1) is herbicide resistance and trait (2) is male or female sterility, whereby male sterility is preferred. In this embodiment, the process of the invention may be used for hybrid seed production for agricultural purposes. Thus, the invention provides a process of producing hybrid seeds, comprising producing a transgenic multi-cellular plant according to the invention (referred to herein as plant A1/A2 in Fig. 13). Preferably, trait (1) is a herbicide resistance and trait (2) is male sterility. Said process of producing hybrid seeds typically further comprises crossing said transgenic multi-cellular plant organism with another plant that is male fertile (referred to herein as plant B in Fig. 13). Plant B should not contain a fragment of a nucleotide sequence encoding said herbicide resistance or said male sterility. The hybrid seeds growing on the male-sterile herbicide resistant plant A1/A2 may then be harvested. The invention also provides the hybrid seed obtained thereby.

The use of said herbicide resistance trait said the process of producing hybrid seeds has the following advantages (cf. Fig. 13): said resistance may be used for selecting plants containing said parent heterologous nucleotide sequence (line A in Fig. 13). Further, said herbicide resistance may be used for selecting male sterile cross-progeny in step (iii) of the invention (cross-progeny of line A1 and line A2 in Fig. 13), as non-sterile self progeny of line A1 and non-sterile self-progeny of line A2 is not herbicide resistant. Consequently, purely male sterile stands of plants may be obtained, and, upon crossing with line B, progeny seeds growing on the male sterile line A1/A2 will be 100% hybrid. Self-progeny seeds growing on plants of line B may be separated by harvesting seeds of line A1/A2 separately from seeds growing on line B. In contrast to prior art processes of producing hybrid seeds using male sterile plant lines, the process of producing hybrid plants disclosed herein is of much more efficiency and less laborious to perform, as the plant lines A1 and A2 may easily maintained by selfing.

Line A containing the pro-locus sequence (Fig. 13) may be male sterile. This is advantageous for generating primary transformants of line A with a desired phenotype (e.g. male sterility, herbicide resistance etc), but maintenance of line A may then be difficult. Line A may therefore be designed such that it is fertile, but lines A1 and A2 may still provide male

sterile plant A1/A2 upon crossing. This may be achieved by separating, in said parent heterologous nucleotide sequence of line A (pro-locus), one of the fragments of the nucleotide sequence encoding the male sterility trait from its promoter. Then, said pro-locus would not provide for male sterility, as one of the fragments encoding male sterility is not expressed. Creation of iso-loci (lines A1 and A2) may bring together promoter and fragment such that said fragment can be expressed, thus allowing to obtain male sterile A1/A2 plants. As an example, said first heterologous nucleotide sequence may interrupt said second heterologous nucleotide sequence in the pro-locus. Upon creation of lines A1 and A2, excision of said first heterologous nucleotide sequence may restore the functionality of said second heterologous nucleotide sequence.

Due to the controlled distribution of both traits to progeny, the cross-progeny (F1 progeny in Fig. 13) will show hybrid vigor and have restored fertility and restored sensitivity to the herbicide the plant A1/A2 was resistant against. Preferably, sterility and herbicide sensitivity are restored in at least 96% of the progeny, more preferably in at least 99% of the progeny. Consequently, said F1 progeny may be used for large scale planting on farm fields without any danger of outcrossing or transferring a functional herbicide resistance gene in the environment.

In example 4 of the invention, engineering of split AHAS gene providing for resistance to imidazoline and sulfonylurea herbicides is described. The AHAS gene was PCR amplified from *Arabidopsis* genomic DNA, mutated and cloned in vectors (Fig. 16) for testing its functionality in transient assays. In example 5, engineering of split barnase providing for a cytotoxic RNase is described. In both examples, we use the intein system to provide for trans-splicing of proteins encoded by split gene fragments. Trans-splicing is mediated by two different intein systems which do not cross-react with each other. This system is based on *Synechocystis* sp. PCC6803 DnaE intein for AHAS and the DnaB intein for barnase. The intermediate constructs with split AHAS-intein fusions and split barnase-intein fusions are shown in Figures 17 and 18, respectively.

Transient test experiments showed the intein-mediated assembly of functional proteins encoded by gene fragments. The invention is not limited to the use of the AHAS gene providing for herbicide resistance. Many other genes conferring herbicide resistance can be used, subject to correct splitting and reconstruction by intein-mediated trans-splicing. Examples of such genes include inter alia 5-enolpyruvylshikimate-3-phosphate synthase, phosphinothricin acetyl

transferase (BAR), betaine aldehyde dehydrogenase (BADH), dihydrofolate reductase (DFR1), acetolactate synthase (ALS), glyphosate oxidoreductase.

Further, barnase is one of several possible genes that may provide for male sterility. Many other genes that affect pollen development when expressed in anther cells or at a desired stage of pollen formation may be employed. Actually, any gene, the gene product of which is capable of interfering with the function and development of pollen can be used in this invention. Examples of such genes inter alia ribosomal inhibitor proteins (Cho *et al.*, 2001, *Mol. Cells*, 11, 326-333), sucrose isomerase (WO159135), protease, glucanase (Tsuchia *et al.*, 1995, *Plant Cell Physiol.*, 36, 487-494), etc. Alternatively, genes responsible for self-incompatibility (preventing self pollination of plants containing said genes) may be used to provide for hybrid seeds production, notably instead of the male sterility trait discussed above (Entani, T., *et al.*, 2003, *Genes Cells*, 8, 203-213; Ushijima, K., *et al.*, 2003, *Plant Cell*, 15, 771-781).

Various pollen or tapetum-specific promoters can be used to drive the expression of a gene/gene fragments for producing male sterility. Examples of tapetum specific promoters are promoters of the A3 and A9 genes (US5723754; Hodge *et al.*, 1991, *J Exp. Botany*, 42, 238 Suppl. p. 46), the tapetum-specific promoter from rice Osg6B gene (Tsuchia *et al.*, 1994, *Plant Mol. Biol.*, 26, 1737-1746), the promoter of tobacco gene TA29 (Kriete *et al.*, 1996, *Plant J.*, 9:809-818), etc. Tissue-specific expression of a gene providing for male-sterility is described in detail in WO98/32325.

In the next step of cloning, said gene fragments were assembled in pairs in intermediate constructs (Fig. 19) designed for final pro-locus vector engineering (Fig. 21) according to the description in example 6. Said pro-locus vector is designed for generation of parental line A, as described in example 8. Said parental line that will be male-sterile can be selected by using the herbicide resistance provided by split AHAS gene. For generating lines A1 and A2 from the parental plant, site-specific recombination may be used. A description of vectors providing for recombinase activity is presented in example 7. The transgenic plants carrying recombinase genes may be generated in the same way as the parental plants carrying pro-locus. Methods of transformation are exemplified in example 8.

In order to generate lines A1 and A2 carrying iso-loci, primary transformants corresponding to the parental line were cross-pollinated with pollen from the plant providing for recombinase activity (example 8). The progeny from such crosses was tested by PCR for the presence of heterologous DNA corresponding to one and the absence of the heterologous DNA

corresponding to the another iso-locus and vice-versa. The generation and structure of iso-loci is shown in Figure 22. Generated lines A1 and A2 carrying different iso-loci were tested for their functionality by cross-pollination. If homozygous lines were used, all progeny from such lines was herbicide resistant and male sterile. In Figure 22, we demonstrate the possibility of generating iso-loci from a pro-locus with the help of one site-specific recombinase. For recombinase PhiC31, recombination (excision or integration) requires two different recombination sites, AttP and AttB. Recombination catalysed by this integrase is an irreversible process, as it leads to the formation of AttL or AttR sites that are not recognised by recombinase PhiC31. The pro-locus shown in Figure 22 contains three such sites and random interaction between two of them (catalysed by the integrase) would lead to excision event with two possible outcomes, generating either line A1 or line A2 with iso-loci. In contrast, a similar approach with parental line transformed with vector pICH12970 (Fig. 21) will produce four different variants of iso-loci with and without HPT selectable marker due to the presence of an additional AttB site.

The approach with said pro-locus in parental line A has important advantages over known hybrid seed production systems: it allows to directly select primary transformants showing the required male sterile phenotype; fertility restoration during the generation of lines A1 and A2 with iso-loci from parental line may be tested. This reduces the time necessary for developing the hybrid seed production system of the invention and makes its maintenance convenient and straightforward.

In addition, the approach of the invention is easily compatible with other methods, for example with methods of controlling seed germination. Controlling seed germination may address specific biosafety issues, especially in the case of producing industrial enzymes, proteins for human and animal health, etc., in hybrid plants. Controlling seed germination can eliminate the problem caused by plant –“volunteers” which frequently contaminate the following harvest and may pose a serious biosafety problem, especially in case of “pharma” proteins. There are several reports addressing the issue of controlling seed germination (US5723765; WO9744465; US5129180; US5977441), however, these methods are not integrated into a process of producing hybrid seeds. Controlling the germination of seeds harvested from hybrid plants may be done according to the general teaching of this invention. Preferably, the hybrid (F1) plant is homozygous for an inactive locus A3 (see Fig. 14D) that can control seed germination after being activated (the activated locus A3 is designated A3* in Fig. 14D). This would provide all progeny of F1 plants with locus A3. Said homozygosity in F1 may be

achieved by introducing a heterologous sequence controlling seed germination in a pre-determined position of a nuclear chromosome of line A1/A2 (or its precursors line A1, A2 or line A) and in line B e.g. via homologous recombination or site-directed integration. Alternatively, introgression of the desired locus by standard breeding methods is also possible. In addition, the hybrid plant (F1) should contain an activator (A4 + B4) for said inactive locus (A3), said activator may be encoded by two heterologous nucleotide sequences, A4 and B4 (Fig. 14D). Sequences A3, A4, and B4 may be brought together as the result of crossing between line A1/A2 and line B to produce F1 plants. In F1 plants, the activator can be rendered functional by intein- or ribozyme-mediated trans-splicing of protein or RNA sequences, respectively, expressed from sequences A4 and B4. Preferably, said activator is a recombinase or a transposase under control of a transiently active promoter (US597741), whereby said promoter is preferably not embryo-, seed- or seed germination specific, i.e. it does not overlap with or precede the expression pattern of the promoter driving the expression of gene(s) of the A3 locus that controls seed germination. The promoter controlling A3 and said gene controlling seed germination (A3) may be separated by a blocking sequence which can be removed by said recombinase/transposase used as said activator. Alternatively, said promoter controlling A3 or said gene controlling seed germination can be re-oriented relative to each other by site-specific recombination, resulting in activation and expression of A3. The activated A3 (A3*) will be inherited to progeny of F1 plants. Self-progeny of F1 plants will be homozygous for A3*, cross progeny of F1 plants will be heterozygous for A3*. Consequently, progeny seeds of the F1 plants will not be viable, i.e. stop growth in an early stage of development.

The development of a plant can be divided into two major groups of stages following germination: vegetative stages (V) and reproductive stages (R). Vegetative stages begin with emergence stage (VE) followed by the cotyledon stage (VC) and by consecutive stages of vegetative development until the beginning of reproductive stages (beginning bloom). Thus, the invention also provides plants grown from the hybrid seeds of the invention, wherein progeny seeds of said (hybrid) plants do not reach the cotyledon stage, preferably they do not reach the VC stage, preferably they do not reach the VE stage, most preferably they do not germinate. Using this embodiment, hybrid plants with a potentially problematic genetic content may be used e.g. for expressing a protein of interest, without the danger that seeds from these plants give rise to unwanted plants in the next growing season.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1

General scheme of intein mediated trans-splicing resulting in functional protein formation.

Figure 2

A - depicts the general principle of the invention, where *trans*-splicing-mediated formation of a functional protein takes place in cells of hybrid progeny;

B – depicts four possible relative locations of the first and the second heterologous nucleotide sequences on host chromosomes of an organism. Case III and IV show relative locations of said heterologous sequences in the transgenic multi-cellular plant of the invention. A diploid organism having two chromosomes and a trait of interest encoded by two fragments (A and B) is used as an example.

C – depicts the basic principle of achieving allelic locations of said first and said second heterologous nucleotide sequences providing for trans-splicing by means of site-targeted integration.

D – depicts the basic principle of achieving allelic locations of said first and said second heterologous DNA sequences providing for trans-splicing by means of transposition.

E – general scheme of methods for achieving allelic locations of different heterologous DNA sequences on homologous chromosomes.

Figure 3 shows schematic representations of T-DNA regions for plasmids pIC5'gfpint and pICintgfp3'.

Figure 4 depicts schematic representations of T-DNA regions for plasmids pIC5'epsp-int and pICint-epsp3'.

Figure 5 depicts schematic representations of T-DNA regions for plasmids pIC5'epsp-intM and pICint-epsp3'M.

Figure 6 depicts a schematic representation of a construct designed for achieving allelic location for the 5' or 3' parts of EPSP coding sequence (A) and its derivatives (B and C)

resulting from excision of non-autonomous transposable elements (*Ds* or *dSpm*, respectively) upon exposure to transposase source.

Figure 7 depicts a general scheme of a construct (center) designed for achieving allelic locations of different heterologous DNA fragments (hDNA 1 and hDNA 2) by way of transposition-mediated removal of unwanted fragments upon exposure to a transposase source. SM – selectable transformation marker; CSM – counter-selectable marker.

On the top, the unwanted fragments excised by the action of the respective transposase are shown. At the bottom, the desired fragment left behind by the transposition are shown.

Figure 8 shows a schematic representation of a method of generating plants with different heterologous DNA fragments (hDNA 1 and hDNA 2) in allelic locations using transposition. A transposase is provided to progeny of plant 1 by crossing plant 1 with plant 2. SM – selectable marker gene; CSM – counter-selectable marker gene; TPase – transposase.

Figure 9 depicts intermediate constructs and Binary vectors used to make constructs shown in figure 3 and 4.

Figure 10 depicts a map of plasmid pICH5300.

Figure 11 depicts a map of Icon Genetics Binary vector pICBV16.

Figure 12 depicts the general schemes for existing genetic/transgenic hybridization systems. Current systems require to engineer three plant lines – a male sterile line, a maintainer line, and a fertility restores line.

Figure 13 depicts schematically the principle of the process of producing hybrid seeds according to the present invention. This system requires to design only one original parental line A with pro-locus containing the parent heterologous nucleotide sequence of the invention. Line A may be herbicide resistant (H^R) and male sterile (ms), allowing selection using the appropriate herbicide for the resistance trait employed. Splitting of said parent heterologous nucleotide sequence leads to line A1 and line A2 containing said first and said second heterologous nucleotide sequence, respectively. Lines A1 and A2 are therefore male fertile and herbicide sensitive (H^S). Lines A1 and A2 may be maintained by selfing. Crossing of line A1 and

line A2 leads to the male sterile and herbicide resistant line A1/A2 of the invention, whereby self-progeny of line A1 and self-progeny of line A2 may be eliminated using said herbicide resistance. Crossing of line A1/A2 with a line B that may be a wild-type (WT) line leads to seeds (F1 progeny) growing on A1/A2 plants. When said F1 progeny seeds are sowed, F1 plants growing therefrom will show hybrid vigor.

Figure 14 A-D shows steps of the process of producing hybrid seeds according to the invention. A - scheme of creating lines A₁ and A₂ with iso-loci from parental line A having a pro-locus containing the parent heterologous nucleotide sequence depicted at the top. Treatment of line A with recombinase A1 removes a part of the parent heterologous nucleotide sequence containing fragments HR5' and MS5', thus forming line A1. Treatment of line A with recombinase A2 removes a part of the parent heterologous nucleotide sequence containing fragments HR3' and MS3', thus forming line A2.

All the gene fragments may be designed as translational fusions with intein fragments capable of trans-splicing. Filled and dotted triangles show the recombination sites recognised by different site-specific recombinases.

SM – selectable marker; HR 3' – 3' fragment of gene conferring herbicide-resistance; HR 5' – 5' fragment of the gene conferring herbicide resistance; MS 3' – 3' fragment of the gene providing for male sterility; MS 5' – 5' fragment of the gene providing for the male sterility.

B - creation of male sterile line (at the bottom in the middle) by crossing line A1 and line A2. Self-progeny of line A1 (left picture at the bottom) and self-progeny of line A2 (right picture at the bottom) can be eliminated due to herbicide sensitivity, allowing pure stands of the male sterile herbicide resistant line A1/A2 (at the bottom in the middle).

C – production of hybrid seeds by crossing line A1/A2 (line A1XA2). All progeny is herbicide sensitive and male sterile. Cross progeny shows hybrid vigor, whereas self-progeny of line B does not. Self-progeny seeds growing on plants of line B may be separated from cross-progeny seeds growing one line A1/A2 by harvesting them separately.

D – shows production of hybrid seeds providing for F2 progeny with controlled seed germination. A3 locus provides for controlling the seed germination once activated (A3*) by activator provided by A4 and B4 loci.

Figure 15 depicts a possible approaches to generate iso-loci. SM – selectable marker. Filled and dotted triangles show the recombination sites recognised by different site-specific

recombinases.

Figure 16 depicts schematic representations of T-DNA regions of plasmids pICH12590 and pICH12600.

Figure 17 depicts a schematic representation of the T-DNA regions of plasmids pICH12610 and pICH12650.

Figure 18 depicts schematic representations of T-DNA regions for plasmids pICH12830 and pICH12840.

Figure 19 depicts a schematic rerepresentation of T-DNA regions of constructs pICH12910 and pICH12950.

Figure 20 depicts schematic representation of T-DNA regions of plasmids pICH12870, pICH13130 and pICH13160.

Figure 21 depicts schematic representation of T-DNA regions of plasmids pICH12960 and pICH12970.

Figure 22 depicts pro-locus from pICH12960 of line A (top) and splitting of the parent heterologous nucleotide sequence for generating iso-loci from a pro-locus of the T-DNA region of pICH12960. The pro-locus contains AttP and AttB recombination sites of an integrase. Application of the integrase leads to statistic removal of of one part of the pro-locus or the other part, thus leading to line A1 and to line A2. Molecular analysis e.g. by PCR is typically be carried out for analysing the recombination result.

DETAILED DESCRIPTION OF THE INVENTION

In this invention, we propose to split the coding sequence of a transgene involved in a trait of interest into two or more fragments that can be bound to each other on the protein level, notably by trans-splicing. Heterologous nucleotide sequences containing these fragments are introduced into the genome of a host plant, preferably into homologous chromosomes, or in the genome and the plastome of a transgenic multi-cellular plant, by hybridising parent plants. Once transcribed and translated, the protein fragments can be assembled by protein trans-splicing, thus forming a functional protein, notably a protein which can provide for the trait of interest. Since the plant breeding process usually involves very specific parental crosses, managing said process of the invention does not pose serious additional problems. Any undesired, spontaneous cross between the transgenic plant of the invention and unwanted organisms effectively disassembles said trait, thus abolishing expression and greatly reducing the chance of functional gene transfer to illicit progeny.

The processes of the invention allow to build mechanisms that would control either the expression of the transgene per se or it could be utilized to control the transgenic variety, as the progeny of any illicit cross is rendered non-viable. Both of these possibilities are *inter alia* contemplated in our invention.

The invention also allows one skilled in the art to design schemes for selecting primary transformants based on a selectable marker that is effective and operable in the T_0 progeny, but fragments or alleles of which, upon subsequent crosses, segregate to different transgenic progeny and thus disappear as a functional selectable marker gene.

Furthermore, the invention allows rapid *in vivo* assembly of different genes by crossing parents that contain different fragments of a transcriptional unit of interest, thus allowing to swap different functional domains, such as translational enhancers, transit or signal or targeting peptides, purification tags, different functional domains of proteins, etc., by simply crossing plants carrying desired fragments of such a functional gene.

There is a description of a hybrid seeds production system based on barnase gene fragments. If said fragments are expressed in the same cell (anther cells), the protein fragments produced associate, whereby barnase activity is restored, generating male sterility (US6392119; Burgess *et al.*, 2002, *Plant J.*, 31, 113-125). Hybrid seeds produced with the help of said approach recover fertility due to the segregation of barnase gene fragments to different gametes, thus causing the inactivation of the cytotoxic gene responsible for male sterility. However, said system has serious limitations as it is built on protein fragment interactions, not

trans-splicing. As the result, said system is temperature-sensitive: temperatures higher than 18°C may restore fertility of the male-sterile line by dissociating the barnase protein fragments.

In the present invention, protein binding and/or trans-splicing can be achieved by using engineered inteins. Inteins were first identified as protein sequences embedded in-frame within protein precursor and excised during protein maturation process (Perler *et al.*, 1994, *Nucleic Acids Res.*, 22, 1125-1127; Perler, F.B., 1998, *Cell*, 92, 1-4). All information and catalytic groups necessary to perform a self-splicing reaction reside in the intein and two flanking amino acids. The chemical mechanism of protein splicing is described in detail by Perler and colleagues (1997, *Curr. Opin. Chem. Biol.*, 1, 292-299) and by Shao & Kent (1997, *Chem. Biol.*, 4, 187-194). Inteins usually consist of N- and C-terminal splicing regions and central homing endonuclease region or small linker region. Over 100 inteins are known so far that are distributed among the nuclear and organellar genomes of different organisms including eukaryotes, archaebacteria and eubacteria (<http://www.neb.com/neb/inteins.html>). It was shown that intein molecules are capable of trans-splicing. The removal of the central homing endonuclease region does not have any effect on intein self-splicing. This also made possible the design of trans-splicing systems, in which the N-terminal and C-terminal fragments of intein are co-expressed as separate fragments and, when fused to exteins (protein fragments, being ligated together with the help of intein), can perform trans-splicing *in vivo* (Shingledecker *et al.*, 1998, *Gene*, 207, 187-195). It was also demonstrated with N- and C- terminal segments of the *Mycobacterium tuberculosis* RecA intein, that protein trans-splicing could take place *in vitro* (Mills *et al.*, 1998, *Proc. Natl. Acad. Sci. USA*, 95, 3543-3548). This phenomenon was also identified for DnaE protein of *Synechocystis* sp. strain PCC6803 (Wu *et al.*, 1998, *Proc. Natl. Acad. Sci. USA*, 95, 9226-9231). Two different genes located more than 700 Kb.p. apart on opposite DNA strands encode this protein. It was also shown that two intein sequences encoded by those genes reconstitute a split mini-intein and are able to mediate protein trans-splicing activity when tested in *Escherichia coli* cells. The intein molecule of the same origin (DnaE intein from *Synechocystis* sp. strain PCC6803) was used to produce functional herbicide-resistant acetolactate synthase II from two unlinked fragments (Sun *et al.*, 2001, *Appl. Environ. Microbiol.*, 67, 1025-29) and 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) (Chen *et al.*, 2001, *Gene*, 263, 39-48) in *E. coli*.

The general principle of intein-mediated trans-splicing is shown in Fig.1.

Yet another well established application of inteins is their use for intein-based protein purification systems (for short review see Amitai & Pietrokovski (1999, *Nature Biotechnol.*, 17,

854-855). The self cleavage of intein from its extein releases extein as free protein molecule after the expose to either pH (Wood *et al.*, 1999, *Nature Biotechnol.*, 17, 889-892) or temperature (Sourthworth *et al.*, 1999, *Biotechniques*, 27, 110-114) changes. Alternatively, nucleophilic agents (e.g. DTT) also initiate cleavage, but such agent remains covalently linked to the released protein (Klabunde *et al.*, 1998, *Nat. Struct. Biol.*, 5, 31-37). To the best of our knowledge, there is no prior art describing the use of intein-mediated protein trans-splicing for assembly of useful traits in plant cells in a biologically safe and controllable way. The general scheme of trans-splicing mediated trait assembly in F_1 progeny is shown in Figure 2A. None of two parental lines (P_1 and P_2) has a fully functional linear gene encoding said trait. In contrast, each contains fragments (A or B) of said gene preferably located on homologous chromosomes. As a result of hybridization between P_1 and P_2 , a progeny is generated that provides for a functional trait due to trans-splicing mediated assembly of proteins encoded by fragments A and B. It is evident from said Figure, that only one fourth of S_1 progeny derived from self-pollination of the primary hybrid will retain the trait of interest, while the other half will inherit only one out of the two fragments required for providing said trait, and one fourth will have neither A or B. It is also evident, that cross-pollination with any other plant (illicit cross) having none of the fragments A and B will not lead to transmission of the trait, as only one of the two fragments necessary for functional gene is transmitted to each progeny plant.

There are several developed approaches and engineered inteins, which can be used to practice this invention (references cited above). They actually cover the use of all known types of inteins in order to engineer trans-splicing events in eukaryotic cells. In EXAMPLE 3 we describe intein-mediated interaction, which brings together two domains of EPSP synthase providing for herbicide resistance. It demonstrates the possibility of assembling a functional protein dimer by bringing together domains necessary for function without actually protein trans-splicing taking place. Such intein-mediated protein-protein interactions also offer an alternative in some specific cases to provide for a trait without protein trans-splicing.

The processes of the invention may be used as a convenient way of assembling a desired sequence and/or expression unit from different parts in *trans*, using as modules or building blocks different transgenic plants. Their hybrid progeny would put together modules inherited from different parents through engineered intein-mediated trans-splicing. It is possible to form a trait of interest by choosing the appropriate pair of transgenic parents containing required modules, very much like by choosing an appropriate pair of parental plants for

producing high value hybrid seeds in traditional breeding. Examples of such modules include different signal peptides, binding domains (e.g. cellulose, pectin, starch binding domains, etc.), retention signals, compartmentalization signals, activation domains, domains with enzymatic activities, affinity tags, regulatory sequences, different genes of interest and parts thereof. In this regard the trait of interest is understood broadly and includes not only a functional protein with a specific capabilities but in particular a protein targeted to a specific compartment or macromolecular matrix, or protein engineered for subsequent isolation/purification.

Additionally, trans-splicing on protein level gives many important advantages which cannot be provided by RNA trans-splicing. Said advantages are the result of the following features:

- a) intein-mediated trans-splicing directly results in the protein molecule, while ribozyme-mediated trans-splicing forms RNA molecule, which, in most cases, shall be translated into the protein, thus restricting the choice of cellular/intercellular compartment for said trans-splicing;
- b) targeting of intein-mediated trans-splicing components provides for a lot of flexibility, as we are dealing with protein molecules, while targeting of RNA molecules is preferably restricted to cytosol;
- c) engineered inteins, in addition to said above, allow for regulating trans-splicing by changing pH, temperature or nucleophilic agents.
- d) Inteins engineered for trans-splicing interact with high efficiency and can bring together protein domains that will provide for enzymatic activity following such interaction even without the covalent link (trans-splicing) taking place.

Such diversity in the choice of parameters for regulation of intein-mediated trans-splicing or interaction (combination of compartmentalization choices with modulation of abiotic parameters) gives flexibility and remarkable variability of choices compared to the RNA-trans-splicing approach.

However, all these potential applications have a limited value without knowing how to achieve the most preferable location of said heterologous fragments relative to each other, preferably on nuclear chromosomes.

According to this invention, said fragments are on different homologous chromosomes (Fig. 2B, case III and IV). In case III, self-progeny can inherit the trait, but said trait will not be inherited by progeny resulting from crossing with plants possessing neither of said fragments if meiotic crossing-over is neglected or absent. Meiotic recombination between the two

homologous chromosomes may, however, physically link said fragments A and B. The frequency of such recombination events directly depends from the relative distance between said fragments on the two homologous chromosomes.

In order to suppress physical linkage of said fragments by meiotic recombination, said fragments are preferably positioned at short relative distance on homologous chromosomes or, most preferably, at the same locus on the homologous chromosomes (Fig. 2B, case IV), thus minimizing the frequency of meiotic recombination between such fragments practically to zero. There are different technical solutions to achieve this most preferable allelic location of said fragments. Said fragments can be integrated at the same locus in pre-engineered integration site by means of site-specific recombination (Fig. 2C). Examples of such systems include the Cre-Lox system from bacteriophage P1 (Austin *et al.*, 1981, *Cell*, 25, 729-736), the Flp-Frt system from *Saccharomyces cerevisiae* (Broach *et al.*, 1982, *Cell*, 29, 227-234), the R-RS system from *Zygosaccharomyces rouxii* (Araki *et al.*, 1985, *J. Mol. Biol.*, 182, 191-203) and the integrase from the *Streptomyces* phage PhiC31 (Thorpe & Smith, 1998, *Proc. Natl. Acad. Sci.*, 95, 5505-5510; Groth *et al.*, 2000, *Proc. Natl. Acad. Sci.*, 97, 5995-6000).

Said fragments A and B can be integrated within chromosomal DNA as one construct AB (Fig. 2D). The design of the construct should allow selective removal of one of the DNA fragments (A or B) using mechanism of controllable DNA rearrangement (excision or transposition), thus generating progeny containing either fragment A or fragment B in the same locus, bringing together both fragments or their transcripts by crossing plants possessing only one of said fragments or both fragments, but only one of required transcripts, will lead to expressing a trait of interest.

An example of said controlled DNA rearrangement is to flank fragments A and B with sequences recognized by different site-specific recombinases and, upon provision of the respective recombinase, to selectively remove either fragment A or fragment B. Alternatively, the placement of a transcription initiation region (promoter) flanked by inverted recombination sites just between said fragments can lead to selective transcription of either fragment A or B depending on said transcription initiation region orientation. The inversion of orientation (but not excision) of said region can be induced by exposure to the recombinase source. As the result, it is possible to achieve selective transcription of either fragment A or fragment B without (physically) removing them, but using DNA inversion as a switch. However, the case of selective expression of one heterologous DNA fragment in the presence of both heterologous

DNA sequences at the same location is not among the preferred embodiments of this invention, as this may not provide the required level of control over trait expression and movement.

An important embodiment of said controlled DNA rearrangement comprises the use of transposition, wherein one of the fragments, for example fragment B, is located and transcribed within a non-autonomous transposable element, and its excision from the construct will trigger transcription of fragment A. Excision of the transposon may or may not be followed by its reinsertion elsewhere and progeny can be selected that contains fragment A or B only. Taking into the consideration that most of transposon reinsertions occur at positions closely linked to the donor site (Jones *et al.*, 1990, *Plant Cell*, 2, 701-707; Carroll *et al.*, 1995, *Genetics*, 139, 407-420), the chance of selecting progeny containing fragments A and B linked in repulsion (on different chromosomes of chromosome pair) is very high. Figure 2E summarizes a variety of approaches for achieving an allelic location of the first and the second heterologous nucleotide sequences including site-targeted integration and excision of said fragments. Transposase-mediated or recombinase-mediated excision of said fragments can be achieved with the help of one (Fig. 2E, a/, a'/, d/ and d'/) or two different transposon or recombinase systems (Fig. 2E, b/, b'/, c/, c'/, e/, e'/, f/ and f'/). The use of two different systems is preferred. The use of two different transposon systems is more preferred. The use of two different transposon systems with overlapping transposon ends is the most preferred (c/ and c'/) embodiment.

The description of construct design for trait assembly through intein-mediated protein trans-splicing (Figures 3, 4) or intein-mediated protein fragment interaction (Fig. 5) is described in examples 1, 2 and 3, respectively. The use of site-specific recombination or transposition allows positioning of the first and the second heterologous nucleotide sequence from a construct at the same loci on homologous chromosomes, which is most favorable for controlled distribution of a trait of interest to cross-progeny. A schematic representation of such a construct (in the T-DNA of a vector for *Agrobacterium*-mediated transformation) and excision of one or the other of said heterologous nucleotide sequences with the use of two different plant transposon systems (*Spm/dSpm* and *Ac/Ds*) is shown in Figure 6. Here, the two components (heterologous nucleotide sequences) of intein trans-splicing system are located on the same T-DNA (Fig. 6 A), but flanked by different transposon ends (*Ds* or *dSpm*) recognized by different transposases, *Ac* or *Spm*, respectively. In brief, the construct in the T-DNA has two non-autonomous transposable elements with overlap at one end. The exposure of a plant or of plant cells carrying such construct to a *Spm* or *Ac* transposase source, will lead to the excision of the

fragment flanked by the *Ds* sequences (exposure to *Ac* transposase), or of the fragment flanked by the *dSpm* sequences (exposure to *Spm* transposase). The resulting T-DNA structures are shown in Fig. 6, B and C, respectively. These resulting constructs are stable even in the presence of *Spm* (in case of B) or *Ac* (in case of C) transposase, as one of the two ends of the non-autonomous transposon required for transposition, is excised together with the other non-autonomous transposable element. Such stabilization of the remaining transposable element is very useful, especially in the case of plants carrying an endogenous transposase source, e.g. corn in case of *Ac* or *Spm* transposase. The scheme of transposon-based selective removal of unwanted DNA fragments is shown in Fig. 7. Here, transposition also leads to removal of other unwanted sequences, e.g. a selectable (SM) and a counter-selectable marker (CSM) genes, thus facilitating the screening for plants/plant cells carrying only the required heterologous nucleotide sequence (hDNA 1 or 2). One of the possible schemes for generating plants with different heterologous nucleotide sequences in allelic relation is shown in Figure 8. These examples with selectable marker genes is not limited to genes conferring antibiotic or herbicide resistance. An extensive list of such genes is shown below. Examples of some counter-selectable marker genes applicable to plant systems, bacterial *codA* and cytochrome P-450 (Kopreck *et al.*, 1999, *Plant J.*, 19, 719-726; Gallego *et al.*, 1999, *Plant Mol Biol.*, 39, 83-93), are described in a number of papers, including their application in combination with transposon systems (Tissier *et al.*, 1999, *Plant Cell*, 11, 1841-1852).

There are well studied transposon systems for plants that are abundantly described in the literature (for reviews see: Dean *et al.*, 1991, *Symp Soc Exp Biol.*, 45, 63-75. Walbot, V., 2000, *Plant Cell Physiol.*, 41, 733-742; Fedoroff, N., 2000, *Proc Natl Acad Sci U S A.*, 97, 7002-7007). The *Ac/Ds* (Briza *et al.*, 1995, *Genetics*, 141, 383-390; Rommens *et al.*, 1992, *Plant Mol Biol.*, 20, 61-70; Sundaresan *et al.*, 1995, *Genes Dev.*, 9, 1797-810; Takumi, S. 1996, *Genome*, 39, 1169-1175; Nakagawa *et al.*, 2000, *Plant Cell Physiol.*, 41, 733-742) and *Spm/dSpm* (Cardon *et al.*, 1993, *Plant J.*, 3, :773-784; Aarts *et al.*, 1995, *Mol Gen Genet.*, 247, 5555-64; Tissier *et al.*, 1999, *Plant Cell*, 11, 1841-1852) systems are well established for transposon tagging in many plant species including many crop plants, and their adoption for practicing this invention is routine for those familiar with the art. This invention is not limited to *Ac/Ds* and *Spm/dSpm* systems. Actually, any transposon system active in plants employing a "cut-and-paste" (excision and reinsertion) mechanism for its transposition can be employed in this invention.

In the examples, we used *Agrobacterium*-mediated T-DNA delivery in plant cells,

whereby said T-DNA contains said first and/or said second heterologous nucleotide sequence as a vector. Different methods may be used for the delivery of vectors into plant cells such as direct introduction of said vector into cells by means of microprojectile bombardment, electroporation or PEG-mediated transformation of protoplasts. *Agrobacterium*-mediated plant transformation is preferred. Thus, DNA may be transformed into plant cells by various suitable technologies such as by a Ti-plasmid vector carried by *Agrobacterium* (US 5,591,616; US 4,940,838; US 5,464,763), particle or microprojectile bombardment (US 05100792; EP 00444882B1; EP 00434616B1). In principle, other plant transformation methods can also be used e.g. microinjection (WO 09209696; WO 09400583A1; EP 175966B1), electroporation (EP00564595B1; EP00290395B1; WO 08706614A1), etc. The choice of the transformation method depends on the plant species to be transformed. For example, microprojectile bombardment may be preferred for monocots transformation, while for dicots, *Agrobacterium*-mediated transformation gives generally better results.

The trans-splicing system described in our invention comprises two fragments, which are provided in trans and are located in allelic positions on homologous chromosomes. This means that our system is better controlled and safer, e.g. it can have zero trait expression level in the uninduced state and zero trait transfer during cross-pollination with other plants.

Genes of interest, or fragments thereof, that can be expressed, in sense or antisense orientation, using this invention include: starch modifying enzymes (starch synthase, starch phosphorylation enzyme, debranching enzyme, starch branching enzyme, starch branching enzyme II, granule bound starch synthase), sucrose phosphate synthase, sucrose phosphorylase, polygalacturonase, polyfructan sucrose, ADP glucose pyrophosphorylase, cyclodextrin glycosyltransferase, fructosyl transferase, glycogen synthase, pectin esterase, aprotinin, avidin, bacterial levansucrase, *E.coli* glgA protein, MAPK4 and orthologues, nitrogen assimilation/metabolism enzyme, glutamine synthase, plant osmotin, 2S albumin, thaumatin, site-specific recombinase/integrase (FLP, Cre, R recombinase, Int, SSVI Integrase R, Integrase phiC31, or an active fragment or variant thereof), isopentenyl transferase, Sca M5 (soybean calmodulin), coleopteran type toxin or an insecticidally active fragment, ubiquitin conjugating enzyme (E2) fusion proteins, enzymes that metabolise lipids, amino acids, sugars, nucleic acids and polysaccharides, superoxide dismutase, inactive proenzyme form of a protease, plant protein toxins, traits altering fiber in fiber producing plants, Coleopteran active toxin from *Bacillus thuringiensis* (Bt2 toxin, insecticidal crystal protein (ICP), CryIC toxin, delta endotoxin, polyopeptide toxin, protoxin etc.), insect specific toxin AaIT, cellulose degrading enzymes, E1

cellulase from *Acidothermus celluloticus*, lignin modifying enzymes, cinnamoyl alcohol dehydrogenase, trehalose-6-phosphate synthase, enzymes of cytokinin metabolic pathway, HMG-CoA reductase, *E. coli* inorganic pyrophosphatase, seed storage protein, *Erwinia herbicola* lycopene synthase, ACC oxidase, pTOM36 encoded protein, phytase, ketohydrolase, acetoacetyl CoA reductase, PHB (polyhydroxybutanoate) synthase, acyl carrier protein, napin, EA9, non-higher plant phytoene synthase, pTOM5 encoded protein, ETR (ethylene receptor), plastidic pyruvate phosphate dikinase, nematode-inducible transmembrane pore protein, trait enhancing photosynthetic or plastid function of the plant cell, stilbene synthase, an enzyme capable of hydroxylating phenols, catechol dioxygenase, catechol 2,3-dioxygenase, chloromuconate cycloisomerase, anthranilate synthase, *Brassica* AGL15 protein, fructose 1,6-biphosphatase (FBPase), AMV RNA3, PVY replicase, PLRV replicase, potyvirus coat protein, CMV coat protein, TMV coat protein, luteovirus replicase, MDMV messenger RNA, mutant geminiviral replicase, *Umbellularia californica* C12:0 preferring acyl-ACP thioesterase, plant C10 or C12:0 preferring acyl-ACP thioesterase, C14:0 preferring acyl-ACP thioesterase (luxD), plant synthase factor A, plant synthase factor B, 6-desaturase, protein having an enzymatic activity in the peroxysomal α -oxidation of fatty acids in plant cells, acyl-CoA oxidase, 3-ketoacyl-CoA thiolase, lipase, maize acetyl-CoA-carboxylase, 5-enolpyruvylshikimate-3-phosphate synthase (EPSP), phosphinothricin acetyl transferase (BAR, PAT), CP4 protein, ACC deaminase, ribozyme, protein having posttranslational cleavage site, protein fusion consisting of a DNA-binding domain of Gal4 transcriptional activator and a transcriptional activation domain, a translational fusion of oleosin protein with protein of interest capable of targeting the fusion protein into the lipid phase, DHPS gene conferring sulfonamide resistance, bacterial nitrilase, 2,4-D monooxygenase, acetolactate synthase or acetohydroxyacid synthase (ALS, AHAS), polygalacturonase, bacterial nitrilase, fusion of amino terminal hydrophobic region of a mature phosphate translocator protein residing in the inner envelope membrane of the plastid with protein of interest to be targeted into said membrane etc.

Any human or animal protein can be expressed, notably in hybrid seeds and plants grown therefrom, using the trans-splicing system of the invention. Examples of such proteins of interest include inter alia the following proteins (pharmaceutical proteins): immune response proteins (monoclonal antibodies, single chain antibodies, T cell receptors etc.), antigens, colony stimulating factors, relaxins, polypeptide hormones, cytokines and their receptors, interferons, growth factors and coagulation factors, enzymatically active lysosomal enzyme, fibrinolytic polypeptides, blood clotting factors, trypsinogen, 1-antitrypsin (AAT), as well as function-

conservative proteins like fusions, mutant versions and synthetic derivatives of the above proteins.

The process of the invention may further comprise expressing a gene encoding a post-transcriptional gene silencing (PTGS) suppressor protein or a function-conservative variant or fragment thereof in a plant for suppressing PTGS of said transgenic coding sequence. Said PTGS suppressor protein gene or function-conservative variant or fragment thereof may be provided to a plant on the same vector carrying said transgenic coding sequence or on an extra vector. Said PTGS suppressor protein is preferably of viral or plant origin. Examples of PTGS suppressor proteins are potato virus X p25 protein, african cassava mosaic virus AC2 protein, rice yellow mottle virus P1 protein, tomato bushy stunt virus 19K protein, rgs CAM or a function-conservative variant or fragment of one of these proteins. Said function-conservative variant or fragment preferably has a sequence identity of 75%, preferably at least 75%, to one of the above proteins. Details on PTGS suppressor proteins and their use can be found in WO0138512.

The invention further provides a transgenic multi-cellular plant organism expressing a trait of interest, said organism having a controlled distribution of said trait to progeny, wherein expression of said trait involves production of a protein molecule by trans-splicing of polypeptide fragments, whereby said polypeptide fragments are encoded on different heterologous nucleotide sequences. Said different heterologous nucleotide molecules are incorporated on homologous chromosomes of this plant. Preferably, said polypeptides form, after trans-splicing or other specific polypeptide interaction, a heterologous protein.

The invention further comprises parts or products of the transgenic plant organisms of the invention and plant seeds obtained by said hybridising. Further, plants or plant material (notably seeds or cell thereof) obtained or obtainable according to step (i) or step (ii) of claim 1. Moreover, vectors for performing the process of the invention are provided, whereby said vectors comprise the parent heterologous nucleotide sequence as defined herein. Further, vectors for performing the process of the invention are provided, notably those shown in the figures and those used in the examples of the invention.

In summary, we propose trait/gene lock systems that are based on a modular principle of providing for trait by assembly of non-functional protein fragments or sub-units into a functional protein. Such systems rely on genetic control of the trait of interest by at least two loci

that segregate independently during crosses, especially during illicit sexual crosses or in the process of a hypothetical horizontal transfer. Based on the present invention, such locks rely on functional protein assembly when the necessary loci are present and expressed in the same cell or in the same organism. Based on our invention, such gain of function is preferably achieved through protein trans-splicing. It was shown before, that intein-mediated trans-splicing allows for functional protein assembly from non-functional protein fragments in vitro (Mills *et al.*, 1998, *Proc. Natl. Acad. Sci. USA*, 95, 3543-3548), as well as in different microorganisms (Shingledecker *et al.*, 1998, *Gene*, 207, 187-195; Wu *et al.*, 1998, *Proc. Natl. Acad. Sci. USA*, 95, 9226-9231; Sun *et al.*, 2001, *Appl. Environ. Microbiol.*, 67, 1025-29; Chen *et al.*, 2001, *Gene*, 263, 39-48). The present invention, however, is not limited to protein trans-splicing as the mechanism of functional protein assembly. Such functionality leading to a trait of interest can be achieved also by providing different subunits of a heteromeric protein, as long as (a) the functionality of the protein of interest depends on the presence of the subunits in questions and (b) the genes for components of are encoded in such a way as to allow for preventing illicit crosses.

The invention also allows to assemble sequence coding for a protein from modules of e.g. signal peptides, binding domains, retention signals, trans-membrane signals, activation domains, domains with enzymatic activities, affinity tags, and regulatory sequences. Such a modular approach makes it simple to find an optimal expression cassette for a specific purpose or for finding an optimal secretory or transit peptide for a specific gene to be over-expressed and accumulated in the cell or a specific compartment thereof. It can be a valuable tool for functional genomics and proteomics studies. A library of plants may e.g. be created, whereby each member of the library contains a particular module (e.g. a specific signal peptide) of one of the above module classes e.g. as said first fragment. The second fragment will then code for a protein of interest. Following said hybridizing, the sequence of said protein is linked to said module by trans-splicing.

Protein splicing, can occur only between at least two genetically designed loci, it occurs in vitro with a very high efficiency, thus allowing for quantitative splicing of parental polypeptides, and it can occur between polypeptides that are encoded in different organellar genomes, such as nuclear genome, plastid or mitochondrial genomes, or extrachromosomal episomes, as long as the translated polypeptides are targeted to the same organelle.

It should be mentioned that the technology described herein can similarly be applied to multi-cellular animals. Humans are excluded.

EXAMPLES

EXAMPLE 1

Intein-mediated trans-splicing of GFP

The 5' end of the GFP gene was amplified by PCR using primers 35spr3 (cgc aca atc cca cta tcc ttc g) and *gfppr8* (ctg ctt gtc ggc cat gat ata g) from plasmid pICH5290 (35S-omega leader-gfp coding sequence-ocs terminator in Icon Genetics binary vector pICBV1 containing BAR for plant selection, Fig. 9). A DNA fragment containing the C-terminal end of the DNAE intein from *Synechocystis* was amplified by PCR from genomic DNA (Strain PCC6803 from the American Type Culture Center) using primers *gfpinte1* (cta tat cat ggcc gac aag cag aag ttt gcgg aatat tgcc tcagt) and *intep2* (ttt gga tcc tta tt aat tgt ccc agc gtc aag). A fusion of the GFP and intein fragments was made by PCR using previously amplified DNA fragments and primers 35Spr3 and *intep2* for the second amplification. The PCR product was cloned as a Nco1-BamHI fragment in pICBV16 (Icon Genetics binary vector with NptII for plant selection Fig. 11; other binary vectors may also be used). The resulting plasmid, pIC5'gfpint (Fig 3), was checked by sequencing.

The 3' end of the GFP gene was amplified by PCR using primers *gfppr9* (aag aac ggc atc aag gtg aac) and *nosterrev* (tca tcg caa gac cgg caa cag g) from plasmid pICH5290. A DNA fragment containing the N-terminal end of the DNAE intein from *Synechocystis* was amplified by PCR from genomic DNA using primers *intep3* (ttt cca tgg tta aag tta tcg gtc gtc) and *integfp2* (gtt cac ctt gat gcc gtt ctt aca att ggc ggc gat cgc ccc att). A fusion of the intein and GFP fragments was made by PCR using previously amplified DNA fragments and primers *intep3* and *nosterrev* for the second amplification. The PCR product was cloned as a Nco1-BamHI fragment in pICBV16. The resulting plasmid, pICintgfp3' (Fig. 3) was checked by sequencing.

The GFP gene product that results from intein-mediated transplicing contains six additional amino acids (KFAEYC) between aminoacids 156 (K) and 157 (Q). This insertion was shown to not significantly affect GFP fluorescence (Ozawa *et al.*, 2001, *Anal. Chem.*, **73**, 5866-5874).

pIC5'gfpint and pICintgfp3' were transformed in agrobacterium strain GV3101 by electroporation. Both constructs were co-expressed transiently in *Nicotiana benthamiana* leaves using *Agrobacterium*-mediated transient expression (Vaquero *et al.*, 1999, *Proc. Natl. Acad. Sci. USA*, 96, 11128-11133). GFP fluorescence was detected when both constructs were co-expressed but not when constructs were expressed individually.

Both constructs were also transformed in *Arabidopsis thaliana* (Col-0) plants as described by Bent *et al.* (1994, *Science*, 285, 1856-1860). Seeds were harvested three weeks after vacuum-infiltration, and germinated and screened for transformants on plates containing 50 mg/L Kanamycin. The same constructs were also used for *Agrobacterium*-mediated leaf disc transformation of *Nicotiana tabacum* plants (Horsh *et al.*, 1985, *Science*, 227, 1229-1231) using 50 mg/L of Kanamycin for selection of transformants. In tobacco and *Arabidopsis*, GFP fluorescence could not be detected in transformants with either construct alone, but was detected in F₁ plants containing both transgenes.

EXAMPLE 2

Intein-mediated trans-splicing of EPSP

The enzyme 5-enolpyruvylshikimate 3-phosphate synthase (EPSP synthase) catalyses the formation of 5-enolpyruvylshikimate 3-phosphate from phosphoenolpyruvate and shikimate 3-phosphate. EPSP synthase is the cellular target of the herbicide glyphosate (N-phosphonomethylglycine). A mutant allele of the *aroA* gene of *Salmonella typhimurium* with a P101S mutation encodes an EPSP synthase with decreased activity to glyphosate. Expression of this gene in plants confers resistance to glyphosate (Comai *et al.*, 1985, *Nature*, 317, 741-744). The 5' end of the mutant EPSP gene was amplified by PCR from *Salmonella typhimurium* genomic DNA (prepared from strain ATCC 39256 from the American Type Culture Center) using primers *epsp1* (tctcc atg gaa tcc ctg acg tta caa c) and *epsprr2* (acc tgg aga gtg ata ctg ttg). A DNA fragment containing the C-terminal end of the DNAE intein from *Synechocystis* was amplified by PCR from pIC5'gfpint using primers *intepr5* (caa cag tat cac tct cca ggt aag ttt gcg gaa tat tgc ctc agt) and *intepr2*. A fusion of the EPSP and intein fragments was made by PCR using previously amplified DNA fragments and primers *epsp1* and *intepr2*. The PCR product was cloned as a NcoI-BamHI fragment in pICH5300 (Icon Genetics binary vector with BAR gene for plant selection, Fig. 10). The resulting plasmid, pIC5'epsp-int (Fig 4), contains the EPSP-N-intein fusion under control of the 35S promoter, fused translationally to an artificial

chloroplast transit peptide (massm lssaav vatra saaqa smvap ftglk saasf pvtrk qnnld itsia snggr vqca). pLC5'epsp-int was checked by sequencing.

The 3' end of the EPSP gene was amplified by PCR from *Salmonella typhimurium* genomic DNA using primers *epsp3* (cgc tat ctg gtc gag ggc gat) and *epsp4* (cgg ggatcc tta ggc agg cgt act cat tc). A DNA fragment containing the N-terminal end of the DNAE intein from *Synechocystis* was amplified by PCR from pLCintgfp3' DNA using primers *intepr3* and *intepr6* (atc gcc ctc gac cag ata gcg gga ttt gtt aaa aca att ggc ggc gat). A fusion of the intein and EPSP fragments was made by PCR using previously amplified DNA fragments and primers *intepr3* and *epsp4*. The PCR product was cloned as a Nco1-BamHI fragment in pCH5300. The resulting plasmid, pLCint-epsp3' (Fig 4), contains the C-intein-EPSP fusion under control of the 35S promoter, fused translationally to the artificial chloroplast transit peptide. pLCint-epsp3' was checked by sequencing.

The EPSP gene product that results from intein-mediated trans-splicing contains ten additional aminoacids (KFAEYCFNKS) between aminoacids 235 (G) and 236 (R). It has been previously shown that this position in the EPSP gene can accommodate insertions of at least 5 to 12 amino acids without compromising gene function (Chen *et al.*, 2001, *Gene*, 263, 39-48).

pLC5'epsp-int and pLCint-epsp3' were transformed in agrobacterium strain GV3101 by electroporation. Both constructs were transformed in *Arabidopsis thaliana* (Col-0) plants as described by Bent *et al.* (1994, *Science*, 285, 1856-1860). Seeds were harvested three weeks after vacuum-infiltration, germinated in soil and screened for transformants by spraying several times with a solution containing 50 mg/L phosphinothricin (PPT).

The same constructs were also used for *Agrobacterium*-mediated leaf disc transformation of *Nicotiana tabacum* plants (Horsh *et al.*, 1985, *Science*, 227, 1229-1231) using 10mg/L of PPT for selection of transformants. Transformants were checked for EPSP gene activity by spraying plants with a commercial formulation of glyphosate (N-phosphonomethylglycine). For both *Arabidopsis* and tobacco, transformants containing either constructs alone did not exhibit glyphosate resistance. F1 plants containing both constructs were resistant to glyphosate.

EXAMPLE 3

Intein-mediated assembly of functional EPSP without trans-splicing

pLC5'epsp-intM is similar to construct pLC5'epsp-int but differ at the junction EPSP-N intein by the addition of 4 native N extein amino acids instead of five and by the first N intein aminoacid

which was changed from Cys to Ala. PIC5'epsp-intM was made following the same strategy as for PIC5'epsp-int except that primer intepr7 (caa cag tat cac tct cca ggt ttt gcg gaa tat gcc ctc agt ttt ggc ac) was used instead of primer intepr5.

PICint-epsp3'M is similar to construct PICint-epsp3' but differ at the junction C intein-EPSP by the addition of 3 C extein amino acids instead of five, the first one mutated from Cys to Ala and the two other native, and by the last C intein aminoacid which was changed from Asn to Ala. PICint-epsp3'M was made following the same strategy as for PICint-epsp3' except that primer intepr8 (atc gcc ctc gac cag ata gcg gtt aaa agc agc ggc ggc gat cgc ccc att g) was used instead of primer intepr6.

The three mutated aminoacids completely prevent intein mediated transplicing but do not prevent association of the N and C intein fragments ((Chen et al., 2001, *Gene*, 263, 39-48). pIC5'epsp-intM and pICint-epsp3'M were transformed in agrobacterium strain GV3101 by electroporation. Both constructs were transformed in *Arabidopsis thaliana* (Col-0) and tobacco as described above. Primary transformants were all sensitive to glyphosate, but hybrid F1 plants containing both constructs, either in tobacco or *Arabidopsis*, exhibited glyphosate resistance.

EXAMPLE 4

Splitting the Arabidopsis AHAS gene

The acetolactate synthase (AHAS) gene from *Arabidopsis* (Genbank accession AY042819) was amplified from *Arabidopsis* genomic DNA using primers *Als1* (5' taaaccatgg cggcggcaac aacaac 3') and *Als2* (5' gactctagac cggtttcac tctcagtatt taatc cggcc atctcc 3') and cloned as an Nco1-Xba1 fragment in Icon Genetics binary vector pICBV24 (Kan^R, selection in *E.coli* and *Agrobacterium*). Ser653 was mutated to Asn by PCR using primers *Alsm5* (5' caggacaagt ctctcgtcgt atg 3'), *Als4* (5' gaaagtgcc ccatcgggga tcatcg 3'), *Als3* (5' cgatgatccc gaatggtggc ac 3') and *Als2*. The amplified mutated fragment was cloned as an Nhe1 - Age1 fragment. A second aminoacid, Pro197 was mutated to Ser by PCR using primers *Als1*, *Alsm5*, *Alsm6* (5' acgacgagag actgtcctg tg 3') and *Alspr1*. The amplified mutated fragment was subcloned as a Sapl-Mlul fragment.

The rice actin1 promoter was amplified by PCR from rice genomic DNA using primers *Actpr1* (5' atgggcgcg cagatctgca tgccggtcga ggctcattcat atgcttgag 3') and *Actpr2* (5' cgccatgggtt

tatcgatagc ttatcgtcta cctacaaaaa agctccgcac g 3'). The PCR product was cloned upstream of the *AHAS* gene as an *Ascl*-*NcoI* fragment. The resulting plasmid, pICH12590 (Fig. 16) contains the rice actin1 promoter followed by the Arabidopsis *AHAS* coding sequence with two mutated aminoacids, and the Nos terminator.

The mutated *Ahas* gene was split into two parts using the *Synechocystis* sp. PCC6803 DnaE intein. To test a position for splitting *AHAS*, aminoacids RAEELLK (aminoacids 428 to 434) were replaced by aminoacids DVKAYCFNKKG using PCR with primers *Alsm5*, *Alsm4* (5' ggccatgggtt aaaacaatat tccgcaaact tgacgtcgtt ctcaagaacc ttattcatcc 3'), *Alsm3* (5' gcggaatatt gttttaacca tggccttgat ttggagttt ggagg 3') and *Nosterrev* (5' tcatcgcaag accggcaaca gg 3'). This substitution results in a protein that is similar to the protein that would be produced by intein-mediated trans-splicing of the constructs described below (pICH12610 and pICH12650, see Fig. 17). The mutated fragment was subcloned as a *BspEI*-*Scal* fragment. The resulting plasmid, pICH12600 (Fig. 16), was tested for *AHAS* activity by bombardment of *Triticum monococcum* cell suspension cultures and selection on plates containing 0.5 to 3 microMolar sulfometuron methyl (Sigma).

The intein-N part of the DnaE intein was amplified by PCR from *Synechocystis* genomic DNA with primers *IntN1* (5' gcaagcttga cgtcaagttt gcggaatatt gcctcagt 3') and *IntN3* (5' cgtctagagt cgacctgcag ttatttaatt gtcccagcgt caag 3'), and subcloned into pICH12600 (Fig. 16) as a *Aat2*-*Xba1* fragment. The resulting plasmid, pICH12610 (Fig. 17), contains the N part of the *AHAS* gene fused to the intein-N fragment.

A PCR fragment containing the Inteín-C part of the DnaE intein was amplified from *Synechocystis* genomic DNA with primers *Ctinte1* (5' ggtctagaatcgatgggttaaagttatcggtcgtcg 3') and *IntC2* (5' cgccatgggtt aaaacaattg gcggcgatcg c 3'). A PCR fragment containing an artificial chloroplast targeting signal (sequence: massmlssaa vvatrassaaq asmvapftgl ksaasfpvtr kqnnlditsi asnggrvqca) was amplified from pICH5300 with primers *Spr3* (5' cgcacaatcc cactatcctt cg 3') and *Ctinte2* (5' cttaaccat agcgcatga actcttcctc c 3'). A fragment containing a fusion chloroplast targeting signal-intein-C fragment was obtained by amplification from both fragments with primers *Spr3* and *IntC2*. This fragment was cloned using *Clal* and *NcoI* into pICH12600 (Fig. 16). The resulting plasmid pICH12650 contains the fusion artificial chloroplast targeting signal- DnaE intein C-AHAS C fragment under control of the rice actin1 promoter, in

a binary vector. To test the functionality of the split *AHAS* gene, pICH12610 and pICH12650 (Fig. 17) were co-bombarded into *Triticum monococcum* cell suspension cultures and the cells selected on media containing 0.5 to 3 microMolar sulfometuron methyl.

EXAMPLE 5

Splitting the *BARNASE* gene

The barnase gene was split using the *Synechocystis* sp. PCC6803 DnaB intein. DNA fragments for the N and C terminal parts of Barnase flanked by appropriate restriction sites were chemically synthesized by a commercial DNA-synthesis company.

The sequence of the N terminal end is:

5' gcaatcgatg gcacagggtta tcaacacggt tgacgggggtt gcggtattatc ttcagacata tcataagcta cctgataatt
acattacaaa atcagaagca caagccctcg gctgggaagt ccgc 3'

The sequence of the C terminal end is:

5' cgccatgggg tggcatcaaa agggaacctt gcagacgtcg ctccggggaa aagcatcggc ggagacatct
tctcaaacag ggaaggcaaa ctccggggca aaagcggacg aacatggcgt gaagcggata ttaactatac atcaggcttc
agaaattcag accggattct ttactcaagc gactggctga ttacaaaac aacggaccat tatcagacct ttacaaaaat
cagataagga tccgc 3'.

The N terminal end of Barnase was fused to the N part of the DnaB intein. The DnaB intein-N fragment was amplified from *Synechocystis* DNA using primers *DnaBintNpr1* (5' gtAAGCTTGA CGTcagagag agtggatgca tcagtggaga tag 3') and *DnaBintNpr2* (5' caCTGCAGct ataattgtaa agaggagctt tctag 3'). The Barnase fragment (a ClaI AatII fragment) and the intein fragment (a AatII PstI fragment) were cloned in an Icon Genetics binary vector resulting in clone pICH12790.

The C terminal end of Barnase was fused to the C part of the DnaB intein. The DnaB intein-C fragment was amplified from *Synechocystis* DNA using primers *dnaBintCpr1* (gt CTG CAG ATC GAT TCA TGA gcc cag aaa tag aaa agt tgt ctc) and *dnaBintCpr2* (tc AAG CTT CCA TGG tct tgc tct tca ctg tta tgg aca atg atg tca t). The intein fragment (a SacI NcoI fragment) and the Barnase fragment (a NcoI BamHI fragment) were cloned in an Icon Genetics binary vector, resulting in clone pICH12820.

Functionality of the N and C terminal Barnase-intein fusion clones was tested by agroinfiltration of *Nicotiana benthamiana* leaves. As expected the infiltrated sector became necrotic.

To reduce activity of Barnase, a frameshift was introduced in the N part of the Barnase gene. A PCR was performed on pICH12790 with primers *Bampr4* (5' gcaatcgatg gcacagggtta ttcaacacgt ttgac 3') that contains the frameshift, and *Bampr5* (5' gcggacgtcc cagccgaggg ctgtgc 3') and subcloned in pICH12790 resulting in plasmid pICH12800. The tapetum-specific promoter (Genbank Number D21160; Tsuchia *et al.*, 1994, *Plant Mol Biol.*, **26**, 1737-1746) was amplified from rice genomic DNA using primers *Tappr1* (5' cggaattcgg cgccttttt ttacacagtt caaagtgaat ttgg 3') and *Tappr2* (5' cgcatcgatg ctttaattagc ttgtgtaat tggag 3') and subcloned in pICH12800 as an EcoRI-ClaI fragment, resulting in plasmid pICH12830 (Fig. 18). The rice tapetum-specific promoter was subcloned from pICH12830 (Fig. 18) into pICH12820 as an EcoRI-ClaI fragment. The resulting construct pICH12840 (Fig. 18) contains the intein-C-Barnase-C fusion under control of the rice tapetum-specific promoter.

EXAMPLE 6

Generation of pro-locus constructs

Assembly of all components required in the final construct was done in a stepwise fashion. First a sequence containing an AttP and an AttB site flanked by appropriate restriction sites was made from overlapping oligonucleotides and cloned in the Icon Genetics binary vector pICBV26 (only contains XhoI-ClaI-XbaI sites between T-DNA borders, Kan^R selection in *E.coli* and *Agrobacterium*). The resulting sequence (agatctgtgc cccaactggg gtaacctttg agttctctc agttgggggc gtagggaatt ctgtctgcag tctagattta tgcattggcgc gcctatctcg agctcgaagc cgcgggtgcgg gtgccagggc gtgcccttgg gctccccggg cgcgtactcc acctaccca tcaactagttg tggtagcatc gcagggccc) is present in construct pICH12920. The N-terminal Barnase-intein fragment (EcoRI-PstI fragment from pICH12830, Fig. 18), the Ahas-intein fragment (AscI XhoI fragment from pICH12610, Fig. 17), and the Ocs terminator (an XbaI PstI fragment from pICH12900) were subcloned in pICH12920. The resulting clone pICH12950 (Fig. 19) contains both N-terminal Barnase and Ahas fragments flanked by AttP and AttB sites in binary vector.

Next, a sequence containing an AttP site flanked by appropriate restriction sites was made from overlapping oligonucleotides and cloned in pICBV26. The resulting construct pICBV12850 contains the sequence (ggtacctgca gtattctaga ttcgaattct cgagtggtggc ggcgcgtgcc ccaactgggg taaccttga gttctctcag ttggggggcgt agggccct) on the T-DNA. The C-terminal intein-Barnase fragment (an EcoRI-BamHI fragment from pICH12840, Fig. 18), the Ocs terminator (a BamHI-Pst1 fragment from pICH12900) and the C-terminal intein-Ahas fragment (an AscI XhoI fragment from pICH12650) were subcloned in pICH12850. The resulting clone pICH12910 (Fig. 19) contains both C-terminal Barnase and Ahas fragments and an AttP site in binary vector.

C-terminal and N-terminal fragments were combined in one binary vector by subcloning a KpnI Apal fragment from pICH12910 into pICH12950 (Fig. 19), resulting in pICH12960 (Fig. 21).

Selectable marker for transformation:

A *HPT* gene under control of the maize ubiquitin promoter was used for plant transformation. To facilitate selection, an intron was inserted into *HPT* coding sequence. First a target site for cloning was inserted into the *HPT* coding sequence by amplifying a PCR fragment from pIC052 (*HPT* coding sequence-Nos terminator in pUC19) with overlapping primers *Bamhpt* (5' cgggatccaa tcagatatga aaaagcctga ac 3'), *Hptint1* (5' ccacaactgt ggtctcaagg tgcttgacat tggggagttc ag 3'), *Hptint2* (5' ggatatcggt ctgttacctc cggaatcggg agcgcg 3'), *Sgfhpt* (5' cgcagcgatc gcatcattg cctccgagac cggctggaga acagcg 3'), and *Inttarg* (5' aggtacgaga ccgatatcca caactgtggt ctcaaggt 3'), and subcloning the amplified fragment as a BamHI SgfI fragment into pIC052. An intron was amplified from petunia genomic DNA with primers *Intpet3* (5' gtctggtctc agtaagttc tgcatttgggt tatgtcctt gcattt 3') and *Intpet4* (5' gtctggtctc tacctgtagc aataattaaa acaaaaata 3') and cloned as a BsaI fragment in the plasmid described above, resulting in plasmid pICH12710. The maize ubiquitin promoter was amplified by PCR from genomic DNA using primers *Ubi1* (5' ttgcatgcct gcagt gcagc gtgacccgggt c 3') and *Ubi2* (5' gggatcctct agagtgcacc tgcagaagta acaccaaaca acaggggtg 3') and cloned as a SphI-BamHI fragment together with *HPT* (a BamHI XbaI fragment from pICH12710) in pICH12720 (an intermediate construct containing restriction sites and an AttB site; sequence 5' tctaagctac tcgagactag tgcattgctgt tctagactcg aagccgcggt gcggtgtgcca gggcggtgcc ttgggtctcc cgggcgcgta ctccacctca ccatcggtgta ccg 3'). The resulting construct pICH12870 (Fig. 20) contains the hygromycin gene with an intron fused to the maize ubiquitin promoter, followed by an AttB site.

Finally, the *HPT* gene was subcloned as a Kpn1-Spe1 fragment into pICH12960 (Fig. 21). The resulting construct pICH12970 (Fig. 21) contains the N and C-terminal ends of Ahas and Barnase fused to intein fragments as well as the HPT selection marker, two AttP sites and two AttB sites.

EXAMPLE 7

Constructing integrase clones

pICH13160 (Fig. 20) was made by cloning the *Streptomyces* Phage C31 integrase (From David Ow, Plant Gene Expression Center, US Department of Agriculture–Agricultural Research Service, Albany, CA 94710, USA) and the *Spm* promoter (amplified by PCR from pIC028 with primers *Spmprfwd* (5' cgtctagagt caaaggagtg tcagtaatt a 3') and *Spmprrev* (5' cgctgcagtg cttggcgagg cgcgcc 3') in an Icon Genetics binary vector (selection in agrobacterium and *E.coli*: Kan^R).

The maize ubiquitin promoter was subcloned from pICH12720 as a BspD1-blunt Pst1 fragment into pICH13160 (Fig. 20) digested with Asc1-blunt and Pst1. The resulting plasmid, pICH13130 (Fig. 9), contains the integrase under control of the maize ubiquitin promoter.

EXAMPLE 8

Generation of transgenic plants with pro-locus

The pICH12970 (Fig.21), pICH13130 (Fig. 20) and pICH13160 (Fig. 20) constructs were transformed into maize, rice and tobacco using Hygromycin selection.

pICH12970 transformants were sprayed with chlorsulfon (Glean, Dupont) to select plants that expressed the mutant split *AHAS* gene at a level sufficient for herbicide resistance (alternatively, construct pICH12960 (Fig. 21) can be transformed into plants using selection on chlorosulfuron or sulfometuron methyl, with the advantage of directly selecting transformants that express *AHAS* at a sufficient level). Plants that looked healthy despite the presence of the split Barnase gene, but that were male sterile, were analyzed by Southern blot to identify individuals containing a single transgene. Such transformants were pollinated by pICH13130 (Fig. 20) or pICH13160 (Fig. 20) transformants. The same transformants were also pollinated with wild type plants to rescue plants with an intact non-recombined transgene locus. The F1

plants (pICH12970 x integrase transformants) were checked by PCR for the presence of both transgenes (pICH12970 transgene, see Fig. 21, and the integrase, see Fig. 20), and seeds were collected. The F2 seedlings were grown and screened by PCR to detect recombinants that lacked either the N-terminal or the C-terminal parts of the split Barnase and Ahas genes. Such plants were completely fertile. Pairs of plants containing complementary parts of the construct (as a result of integrase-mediated recombination) were crossed. Seedlings obtained from these crosses were sprayed with Chlorsulfuron to eliminate plants that did not contain both parts of the construct. All plants that were resistant to chlorsulfuron were also male sterile.

The following methods were used for generating transgenic plants:

Tobacco transformation

The constructs were used for *Agrobacterium*-mediated leaf disc transformation of *Nicotiana tabacum* plants (Horsh *et al.*, 1985, *Science*, 227: 1229-1231) using selection on Hygromycin (25-100 mg/l) or sulfometuron methyl (0.5- 3.0 microM) or chlorsulfuron (0.2-5.0 microM).

Rice transformation

Callus cultures were induced from mature and immature embryos of rice cvs. Pusa Basmati 1, Koshihikari etc.

The culture media were based on Chu (N6) salts and vitamins (Chu *et al.*, *Scientia Sinica*, 18(5):659-68, 1975).

Callus induction and propagation medium was supplemented with 30 g/l sucrose, 600 mg/l L-proline, 2.0 mg/l of 2,4-D and 0.3% gelrite.

Pre-regeneration medium contained N6 salts and vitamins with 30 g/l sucrose, 1 mg/l NAA, 2 mg/l BA, 2 mg/l ABA and 0.6% gelrite.

Regeneration medium contained N6 salts and vitamins with 30 g/l sucrose, 0.2 mg/l NAA, 2 mg/l BA, and 0.6% gelrite.

Infection medium (IM) contained N6 salts and vitamins with 2 mg/l 2,4-D, 10 g/l glucose, 60 g/l maltose, 50 mg/l ascorbic acid, 1 g/l MES (2-N-morpholinoethanesulfonic acid) and 40 mg/l Acetosyringone (AS). The pH of the medium was adjusted to 5.2 by 1 N KOH.

Cocultivation medium (CM) was same as the IM (excluding ascorbic acid) and was solidified by adding 0.6% gelrite.

Infection medium was filter sterilized, whereas all other media were autoclaved. AS, dissolved in DMSO (400 mg/ml), was added after sterilization.

Agrobacterial cultures (strains AGL1, EHA105, LBA4404 etc.) with the appropriate binary plasmids were grown for 3 days at room temperature on LB2N (LB medium with 2 g/l NaCl and 1.5% agar) plates supplemented with the appropriate antibiotics. Bacteria were scraped from the plates and resuspended in the IM (10-20 ml) in 50-mL falcon tubes. The tubes were fixed horizontally to a shaker platform and shaken at low speed for 4 to 5 h at room temperature. Optical density of the bacterial suspension was measured and OD600 was adjusted to 1.0-2.0.

Callus pieces were incubated in the agrobacterial suspension for 20-180 min at room temperature, blotted on the filter paper disks and transferred to the gelrite-solidified CM with 60 g/l maltose. After 3-6 days of cultivation on the CM the calli were washed five times by sterile water and transferred to the gelrite-solidified CM with 60 g/l sucrose and appropriate selective agent and, if needed, 150 mg/l Timentin.

Resistant calli developed under selection were plated to the pre-regeneration medium with appropriate selective agent. Two weeks later the cultures were transferred to the regeneration medium with appropriate selective agent. Regenerated plantlets were grown on half-strength N6 medium without hormones for one month before transplanting into the soil.

Hygromycin B, for hpt (hygromycin phosphotransferase) gene-based selection, was used at concentrations 25-100 mg/l. Selection based on the herbicide-resistant forms of AHAS (Acetohydroxy acid synthase) gene was performed on sulfometuron methyl (0.5- 3.0 microM) or chlorsulfuron (0.2-5.0 microM).

Maize transformation

Maize immature embryos and callus cultures obtained from the lines A188, Hill etc. were transformed essentially in the same way as rice cultures. Most of the media and transformation steps were the same. Pre-regeneration medium was not used. Regeneration medium contained N6 salts and vitamins, 30 g/l sucrose, 2 mg/l Zeatin and 0.05 mg/l 2,4-D. Silver thiosulfate was included in the regeneration medium at concentrations 0.01-0.06 mM.

Table 1. Comparative analysis of ICON hybrid seed production system

Parameters	ICON	Mechanical	Chemical	Genetic	Transgenic
Universality	YES	NO	NO	NO	YES
Female line production	YES	YES	YES	YES	YES/NO
Non-leakiness	YES	YES	NO	NO	NO
Full fertility restoration	YES	YES	YES	YES/NO	NO
Biological safety	YES	YES	YES	YES	NO
Compatibility with other systems	YES	NO	YES	NO	YES/NO
Integrated solution	YES	NO	NO	YES	NO

Claims

1. A process of producing a transgenic multi-cellular plants or parts thereof expressing a trait of interest, said trait having a controlled distribution of said trait to progeny, wherein said process comprises
 - (i) producing a first plant or a cell thereof having in a first locus of a nuclear chromosome a first heterologous nucleotide sequence comprising a first fragment of a nucleotide sequence encoding said trait of interest,
 - (ii) producing a second plant or a cell thereof having in a second locus of a nuclear chromosome homologous to said nuclear chromosome of step (i), a second heterologous nucleotide sequence comprising a second fragment of the nucleotide sequence encoding said trait of interest, and
 - (iii) hybridising said first and said second plant or cells thereof to generate progeny exhibiting said functional trait of interest due to binding between a protein or polypeptide encoded by said first heterologous nucleotide sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence.
2. The process of claim 1, wherein said multi-cellular plant organisms or said parts express two traits of interest, a trait (1) and a trait (2), both traits having a controlled distribution to progeny.
3. The process of claim 2, whereby
 - (i') said first heterologous nucleotide sequence of step (i) comprises:
a first fragment of a nucleotide sequence encoding trait (1) and
a first fragment of a nucleotide sequence encoding trait (2); and
 - (ii') said second heterologous nucleotide sequence of step (ii) comprises:
a second fragment of a nucleotide sequence encoding trait (1) and
a second fragment of a nucleotide sequence encoding trait (2); and
 - (iii') step (iii) comprises hybridising said first and said second plant or cells thereof to generate progeny exhibiting trait (1) and trait (2), whereby exhibiting of trait (1) is due to binding between a protein or polypeptide encoded by said first heterologous nucleotide

sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence.

4. The process of claim 3, wherein the progeny generated in step (iii') exhibits trait (2) due to binding between a protein or polypeptide encoded by said first heterologous nucleotide sequence and a protein or polypeptide encoded by said second heterologous nucleotide sequence.
5. The process of one of claims 2 to 4, wherein said progeny exhibits trait (1) and/or trait (2) due to intein-mediated trans-splicing.
6. The process of claim 2, wherein exhibiting of trait (1) and/or of trait (2) is due to RNA trans-splicing of an RNA expression product encoded by said first heterologous nucleotide sequence and an RNA expression product encoded by said second heterologous nucleotide sequence.
7. The process of one of claim 2 to 6, wherein step (iii) involves selecting progeny that exhibits said trait (1) and said trait (2).
8. The process of one of claims 2 to 7, wherein trait (1) is a herbicide resistance.
9. The process of claim 8, wherein step (iii) involves selecting progeny that exhibits said trait (2) by applying a herbicide to said progeny, whereby said trait (1) endows resistance against said herbicide.
10. The process of one of claims 2 to 9, wherein trait (2) is male or female sterility.
11. The process of one of claims 9 or 10, wherein step (iii) involves selecting progeny that exhibits male sterility as said trait (2) by applying a herbicide to said progeny, whereby said trait (1) endows resistance against said herbicide.
12. The process of one of claims 1 to 11, wherein two or more traits are assembled in step (iii) by trans-splicing.

13. The process of one of claims 1 to 12, wherein steps (i) comprises introducing said first heterologous nucleotide sequences into said first locus of a nuclear chromosome of a plant or a plant cell by site-targeted integration into a pre-engineered integration site or by homologous recombination.
14. The process of one of claims 1 to 13, wherein steps (ii) comprises introducing said second heterologous nucleotide sequences into said second locus of a nuclear chromosome of a plant or a plant cell by site-targeted integration into a pre-engineered integration site or by homologous recombination.
15. The process of one of claims 1 to 14, wherein steps (i) and (ii) are carried out by
 - (a) introducing a parent heterologous nucleotide sequence comprising said first and said second heterologous nucleotide sequences into a nuclear chromosome of parent organisms or cells thereof,
 - (b) optionally selecting organisms or cells thereof having said parent heterologous nucleotide sequence integrated in a desired chromosome or chromosome locus,
 - (c) subsequently splitting said parent heterologous nucleotide sequence so that said first and said second heterologous nucleotide sequences are located on homologous chromosomes in different plant organisms or cells.
16. The process of claim 15, wherein step (a) is carried out by homologous recombination or by site-targeted integration of said parent heterologous nucleotide sequence into a predetermined locus of a nuclear chromosome.
17. The process of claim 15 or 16, wherein step (a) or (b) are followed by producing organisms or cells thereof which are homozygous for said parent nucleotide sequence.
18. The process of claim 15 or 16, wherein said organism obtained in step (a) or step (b) is heterozygous for said parent nucleotide sequence.

19. The process of one of claims 15 to 18, wherein step (c) comprises excision of said second heterologous nucleotide sequence from said parent heterologous nucleotide sequence,
optionally followed by reintegration of said excised second heterologous nucleotide sequence into a locus of a chromosome that is homologous with respect to the chromosome of said parent heterologous nucleotide sequence.
20. The process of one of claims 15 to 18, wherein step (c) comprises excision of said first heterologous nucleotide sequence from said parent heterologous nucleotide sequence, optionally followed by reintegration of said excised first heterologous nucleotide sequence into a locus of a chromosome that is homologous with respect to the chromosome of said parent heterologous nucleotide sequence.
21. The process of one of claims 19 or 20, wherein the plants or cells thereof obtained in claim 19 or 20 or progeny thereof are analysed for said reintegration of the excised heterologous nucleotide sequence,
and plants or cells thereof are selected that do not contain said excised heterologous nucleotide sequence or that contain said heterologous nucleotide sequence at a desired locus on a chromosome homologous to the chromosome harboring the heterologous nucleotide sequence that has not been excised.
22. The process of one of claims 19 to 21, wherein said first and/or said second heterologous nucleotide sequence in said parent heterologous nucleotide sequence is/are contained in a non-autonomous transposon and said excision comprises providing a transposase for said transposon.
23. The process of claim 22, wherein
 - (A) said first heterologous nucleotide sequence in said parent heterologous nucleotide sequence is contained in a first non-autonomous transposon and said second heterologous nucleotide sequence is contained in a second non-autonomous transposon and
 - (B) said first heterologous nucleotide sequence is excised by providing a first transposase functional with said first non-autonomous transposon and said second

heterologous nucleotide sequence is excised by providing a second transposase functional with said second non-autonomous transposon.

24. The process of claim 23, wherein said first and said second transposons in said parent heterologous nucleotide sequence overlap such that excision of said first or said second heterologous nucleotide sequence leads to disruption of said second or said first non-autonomous transposon, respectively.
25. The process of one of claims 19 to 21, wherein said first heterologous nucleotide sequence in said parent heterologous nucleotide sequence is flanked by recombination sites of a first site-specific recombinase and wherein said second heterologous nucleotide sequence in said parent heterologous nucleotide sequence is flanked by recombination sites of a second site specific recombinase.
26. The process of claim 25, wherein said first site-specific recombinase is different from said second site-specific recombinase.
27. The process of claim 25, wherein a segment 1 and a segment 2 of said parental heterologous nucleotide sequence overlap, whereby segment 1 comprises said first heterologous nucleotide sequence flanked by the recombination sites functional with said first site-specific recombinase and segment 2 comprises said second heterologous nucleotide sequence flanked by the recombination sites functional with said second site-specific recombinase.
28. The process of one of claims 15 to 21, wherein said first heterologous nucleotide sequence in said parent heterologous nucleotide sequence is flanked by differing recombination sites of a site-specific integrase and said second heterologous nucleotide sequence in said parent heterologous nucleotide sequence is flanked by differing recombination sites of the same site-specific integrase, and step (c) is carried out by
 - providing said site-specific integrase to said parent organism or cells thereof,

- selecting progeny of said parent organism or cells thereof containing said first heterologous nucleotide sequence but not said second heterologous nucleotide sequence, and
 - selecting progeny of said parent organism or cells thereof containing said second heterologous nucleotide sequence but not said first heterologous nucleotide sequence.
29. The process of one of claims 22 to 28, wherein said transposase, said site-specific recombinase, and said site-specific integrase is provided by hybridising or crossing with a plant or plant cells containing a gene coding for said transposase or said recombinase or by *Agrobacterium*-mediated transformation, viral transfection, particle bombardment, electroporation or PEG-mediated transformation with a gene coding for said transposase or said recombinase.
30. The process of one of claims 1 to 29, whereby said first and said second loci are selected for a reduced probability of undergoing crossing over.
31. The process of claim 30, wherein said first and said second loci are corresponding loci on said homologous chromosomes.
32. The process of one of claims 1 to 31, wherein said first and said second plant or cells thereof are made homozygous for said first and said second heterologous nucleotide sequences.
33. The process of one of claims 1 to 32, wherein said binding of said proteins or polypeptides is followed by peptide bond formation between said protein or polypeptides.
34. The process of claim 33, wherein said binding and said peptide bond formation is intein-mediated trans-splicing.
35. The process of one of claims 1 to 34, wherein said controlled distribution means that, upon crossing of said transgenic multi-cellular plant organism with an organism devoid

of said first and said second heterologous sequences, the frequency of the appearance of said trait in descendent organisms is less than 1%, preferably, less than 0.1%, more preferably less than 0.01%, most preferably less than 0.001%.

36. The process of claim 35, wherein said transgenic multi-cellular plant organism is incapable of expressing said trait of interest in the absence of either said first or said second heterologous nucleotide sequence.
37. The process of one of claims 1 to 36, wherein said multi-cellular plant is further genetically or transiently modified for providing functions necessary for said binding and/or said peptide bond formation.
38. The process of one of claims 1 to 37, wherein said first and/or said second heterologous nucleotide sequence contains an intron for RNA cis-splicing of a transcription product of said first or said second heterologous nucleotide sequence.
39. The process of one of claims 1 to 38, wherein said trait of interest is involved in male or female sterility.
40. The process of one of claims 1 to 39, wherein said trait is selected from the following group: herbicide resistance, insecticide resistance, selectable marker, counter-selectable marker, transcription factor, DNA or RNA modifying enzymes, production of a protein of interest.
41. The process of one of claims 1 to 40, wherein said multi-cellular plant organism is capable of producing progeny.
42. The process of one of claims 1 to 41, wherein said binding and/or said peptide bond formation generates a protein having a polypeptide linked thereto, whereby said polypeptide is selected from the following group: signalling, targeting, and membrane transduction polypeptide; a binding domain, a recognition or a visualisation tag, a purification tag, a protein cleavage sequence.

43. The process of one of claims 1 to 42, wherein said second and/or said first fragment of said gene encoding said trait of interest is operably linked to a regulated promoter.
44. The process of one of claims 1 to 43, wherein said first or said second fragment of said gene encoding said trait of interest encodes an internal ribosome entry site (IRES) allowing translation of a transcript of said first or second fragment.
45. Plant, seed or plant cell expressing a trait of interest, obtained or obtainable according to one of claims 1 to 44, and products derived therefrom.
46. Process of producing hybrid seeds, comprising producing a transgenic multi-cellular plant according to one of claims 2 to 44.
47. The process of producing hybrid seeds according to claim 46, further comprising crossing said transgenic multi-cellular plant organism with another plant that is male fertile.
48. The process of producing hybrid seeds according to claim 46 or 47, wherein trait (1) is a herbicide resistance and trait (2) is male sterility.
49. The process of one of claims 46 to 48, wherein progeny seeds of plants that were grown from said hybrid seeds do not reach the cotyledon stage.
50. The process of claim 49, wherein progeny seeds of plants that were grown from said hybrid seeds do not germinate.
51. The process of one of claims 48 to 50, whereby said transgenic multi-cellular plant organism and/or said other plant contain(s) a non-expressable seed germination control gene that is rendered expressable in plants grown from said hybrid seeds.
52. Hybrid seeds obtained or obtainable according to the process of one of claims 46 to 51 and plants grown therefrom.

53. Plants grown from the hybrid seeds of claim 52, wherein the progeny of said plants is non-viable.
54. The plants according to claim 53, wherein progeny seeds of said plants do not germinate.
55. The plants of claim 53 or 54, containing an inactive seed germination control gene that can be activated by expressing an activating protein.
56. Use of the hybrid seeds or plants according to one of claims 52 to 55 for expressing a protein of interest, notably a pharmaceutical protein..
57. Plant or seed or cell thereof obtained or obtainable according to step (i) or step (ii) of one of claims 1 to 44.

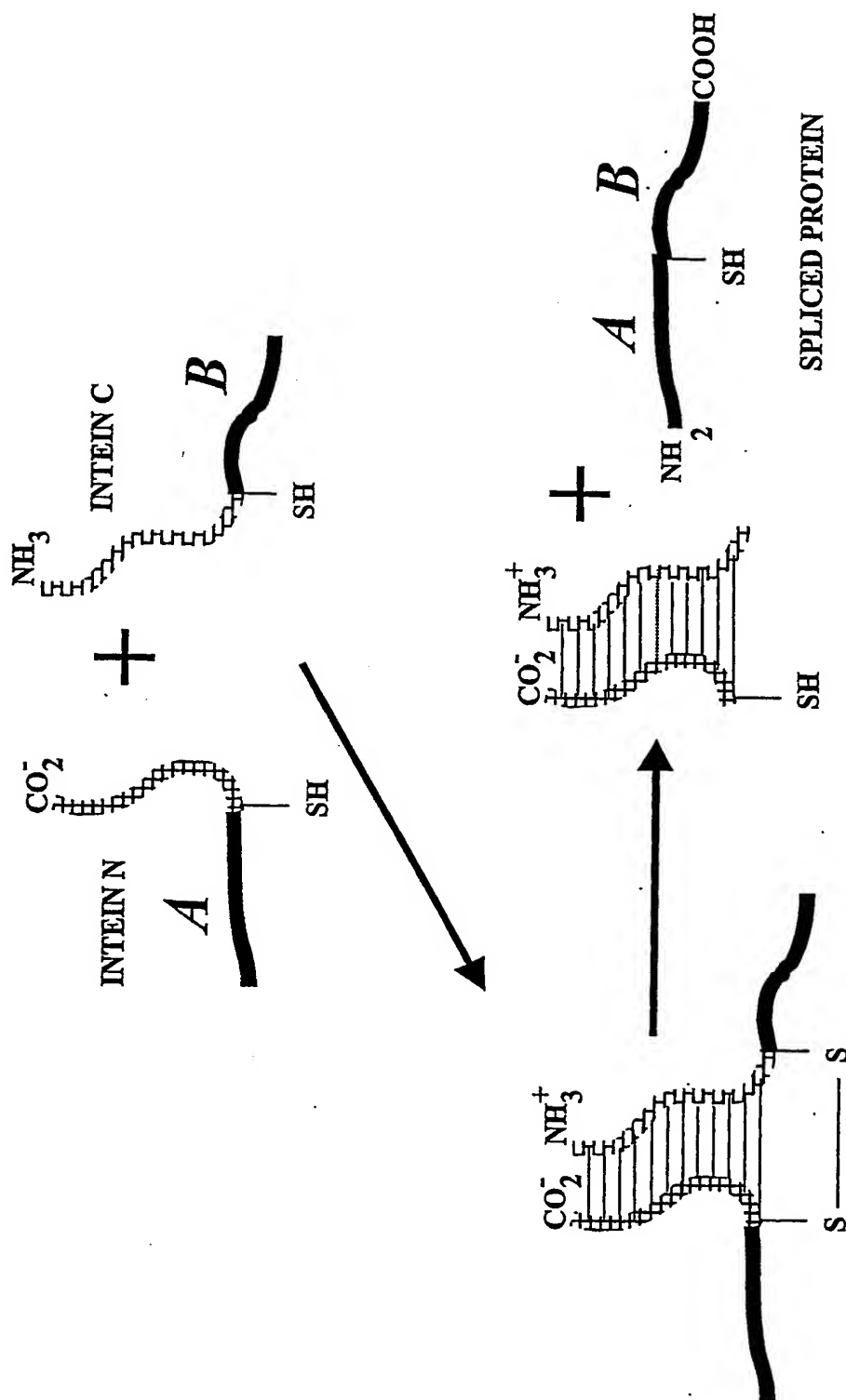


Figure 1

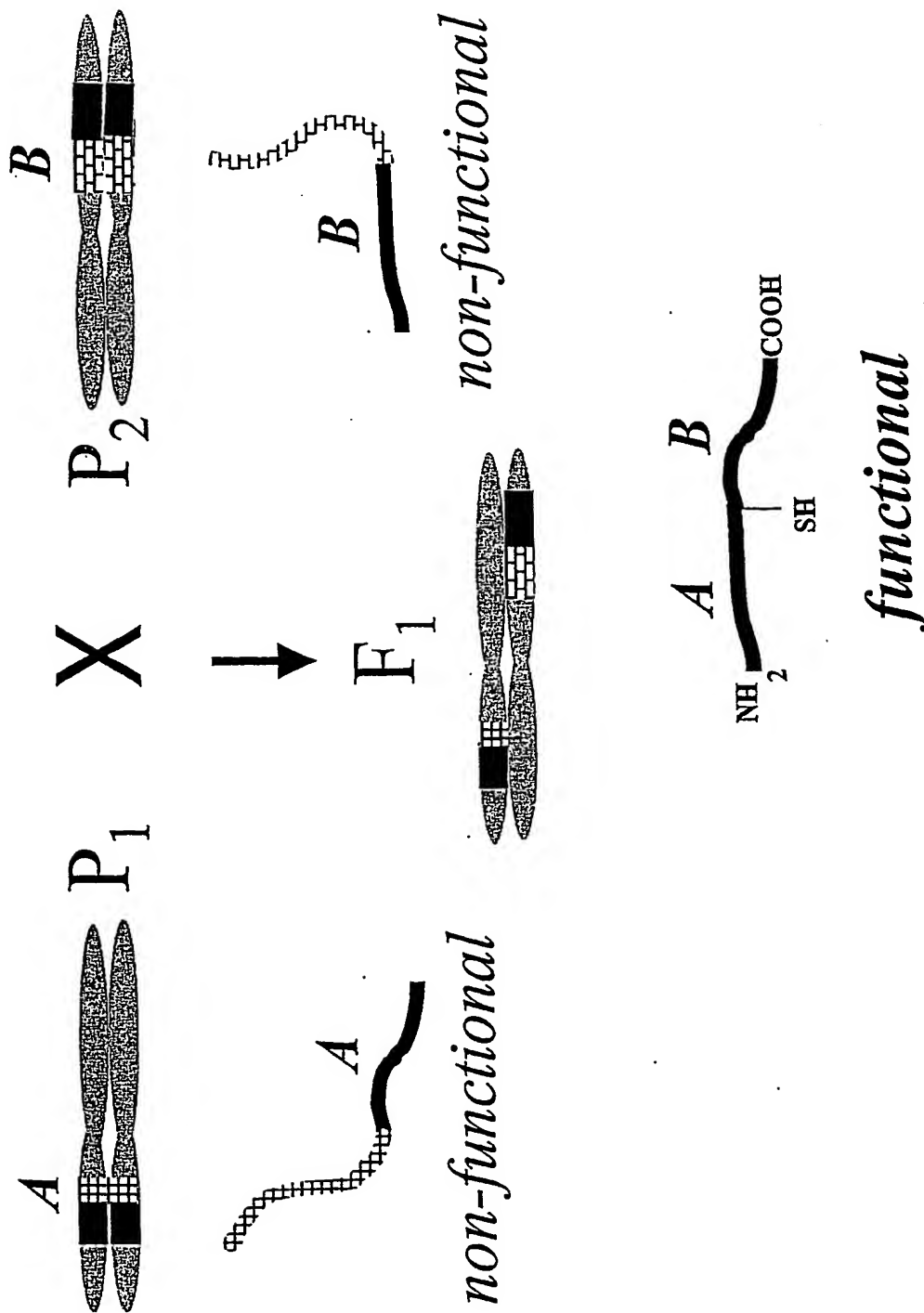


Figure 2A

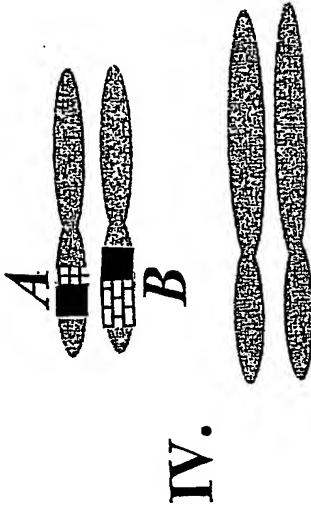
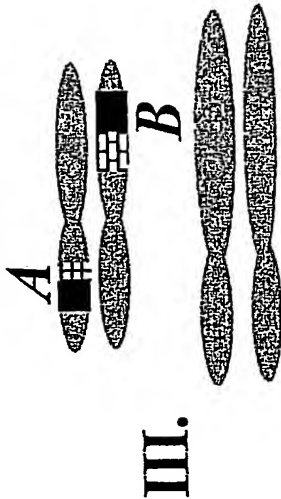


Figure 2B

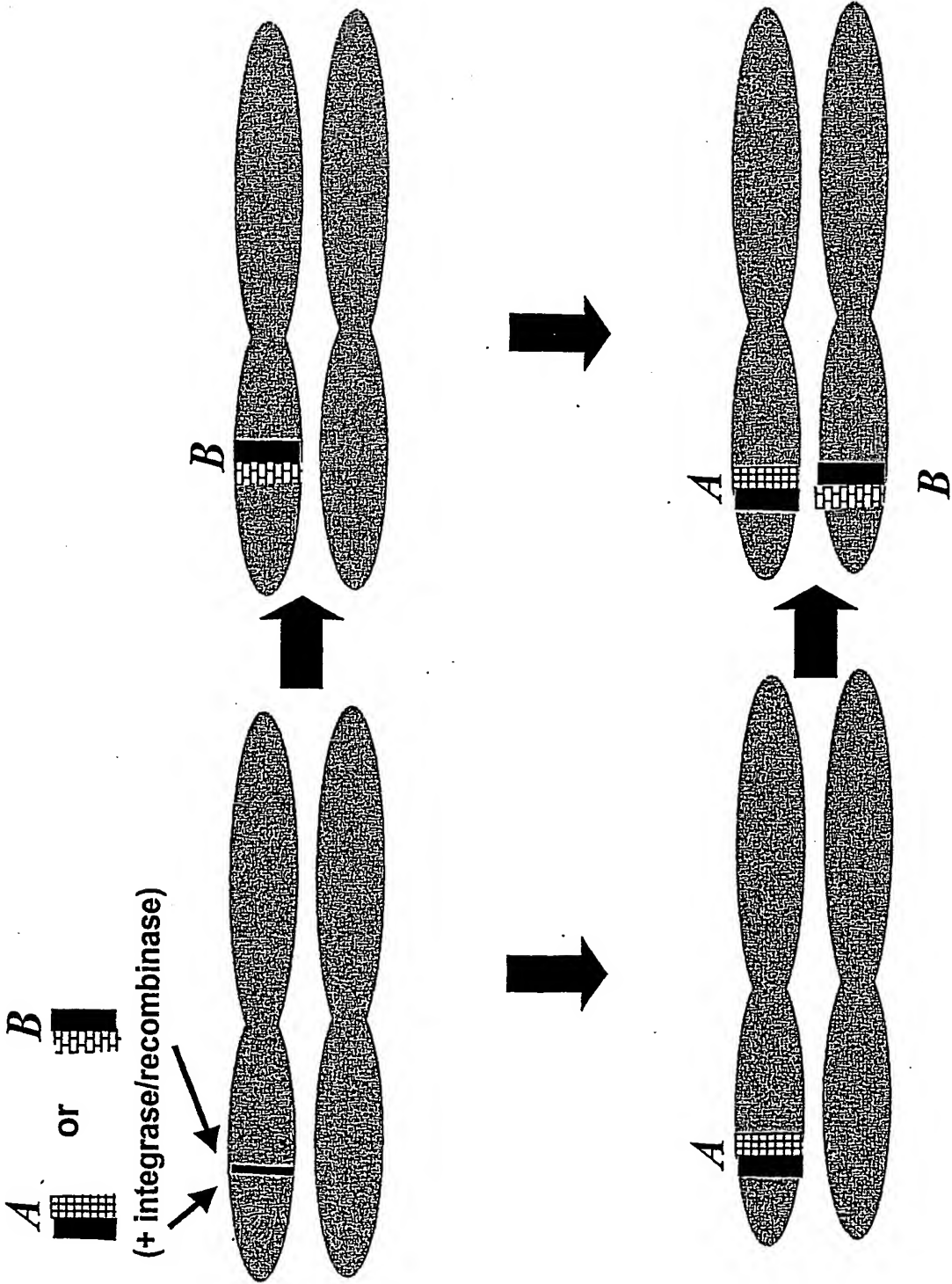


Figure 2C

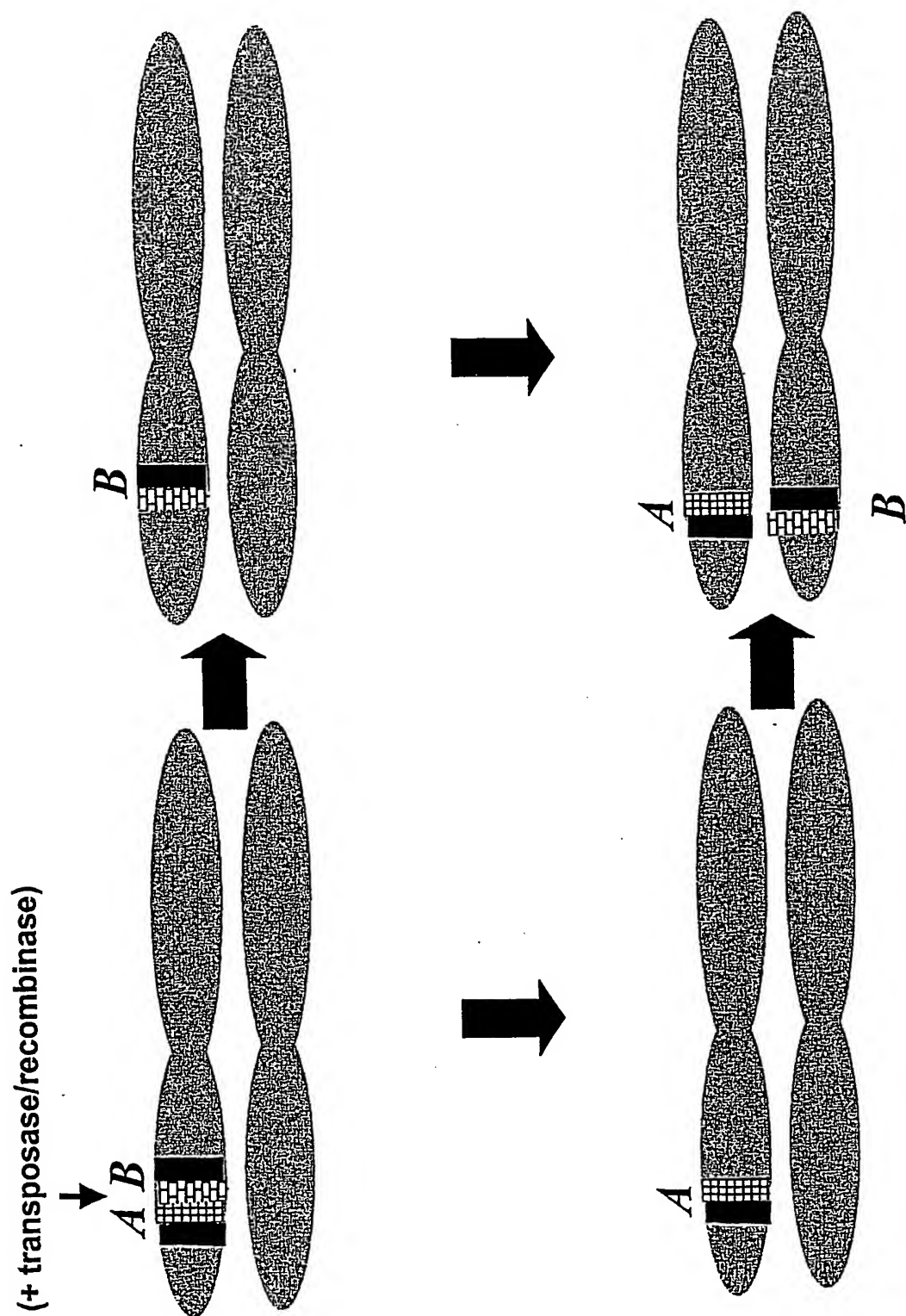


Figure 2D

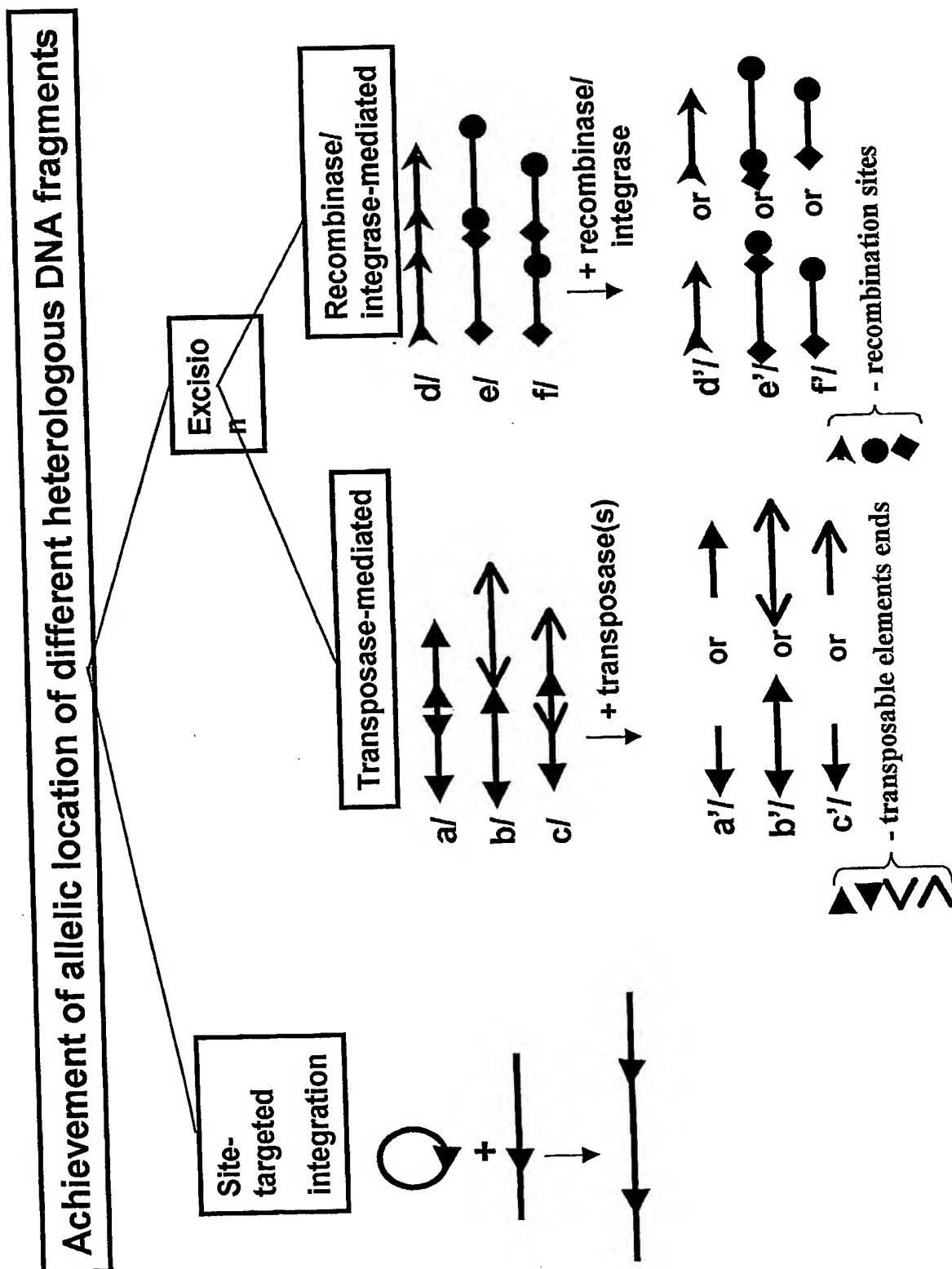


Fig. 2E

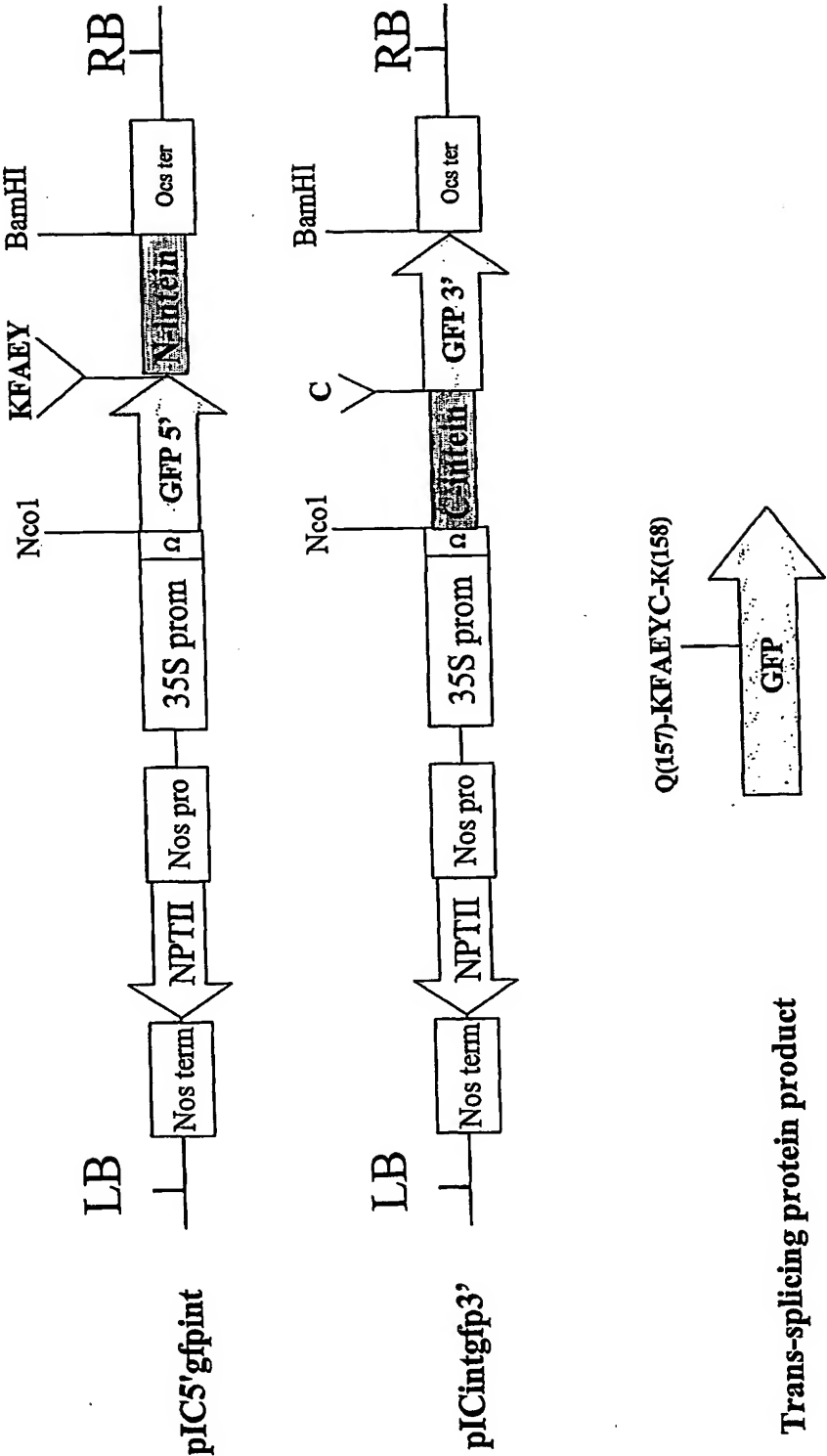


Fig. 3

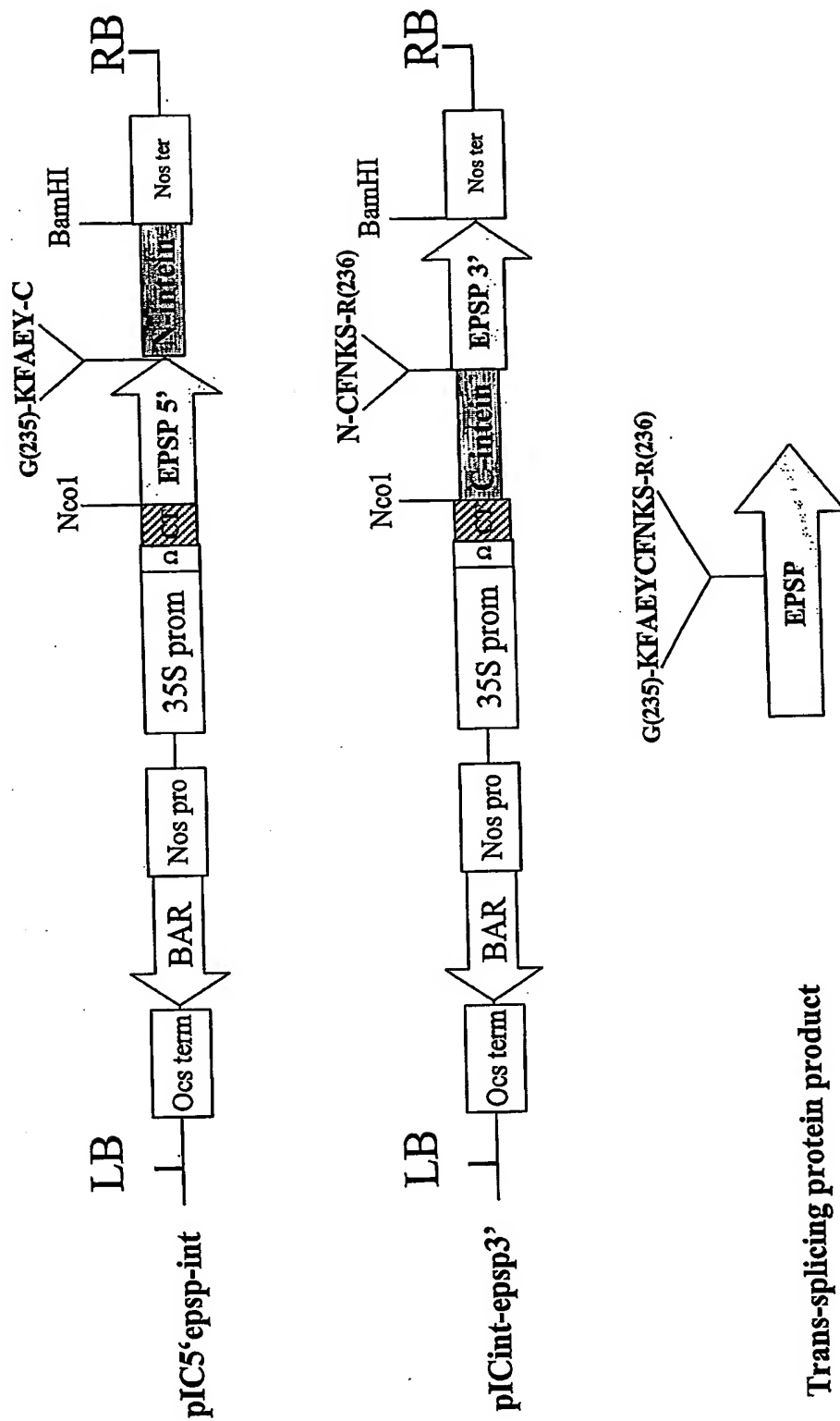


Fig. 4

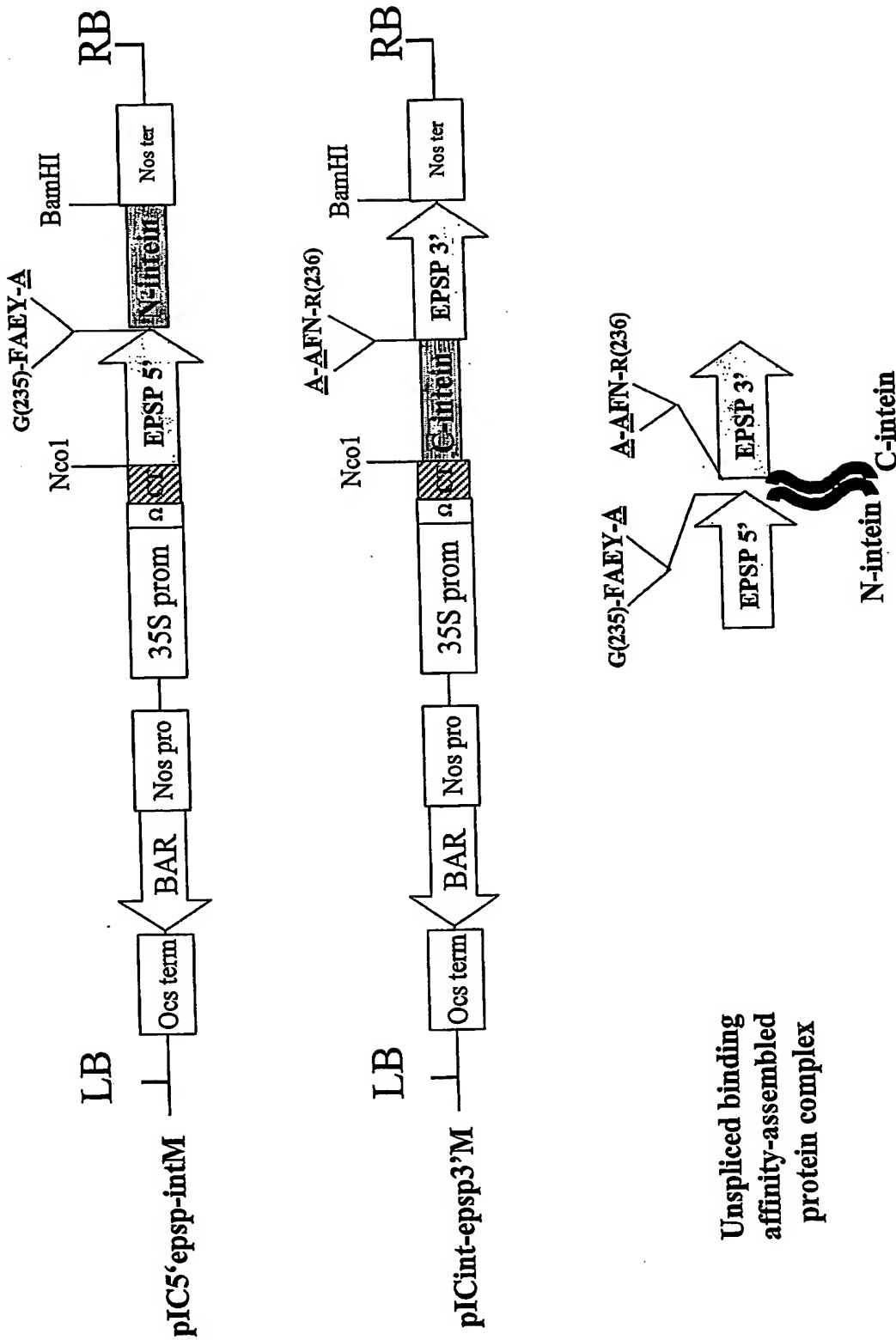


Fig. 5

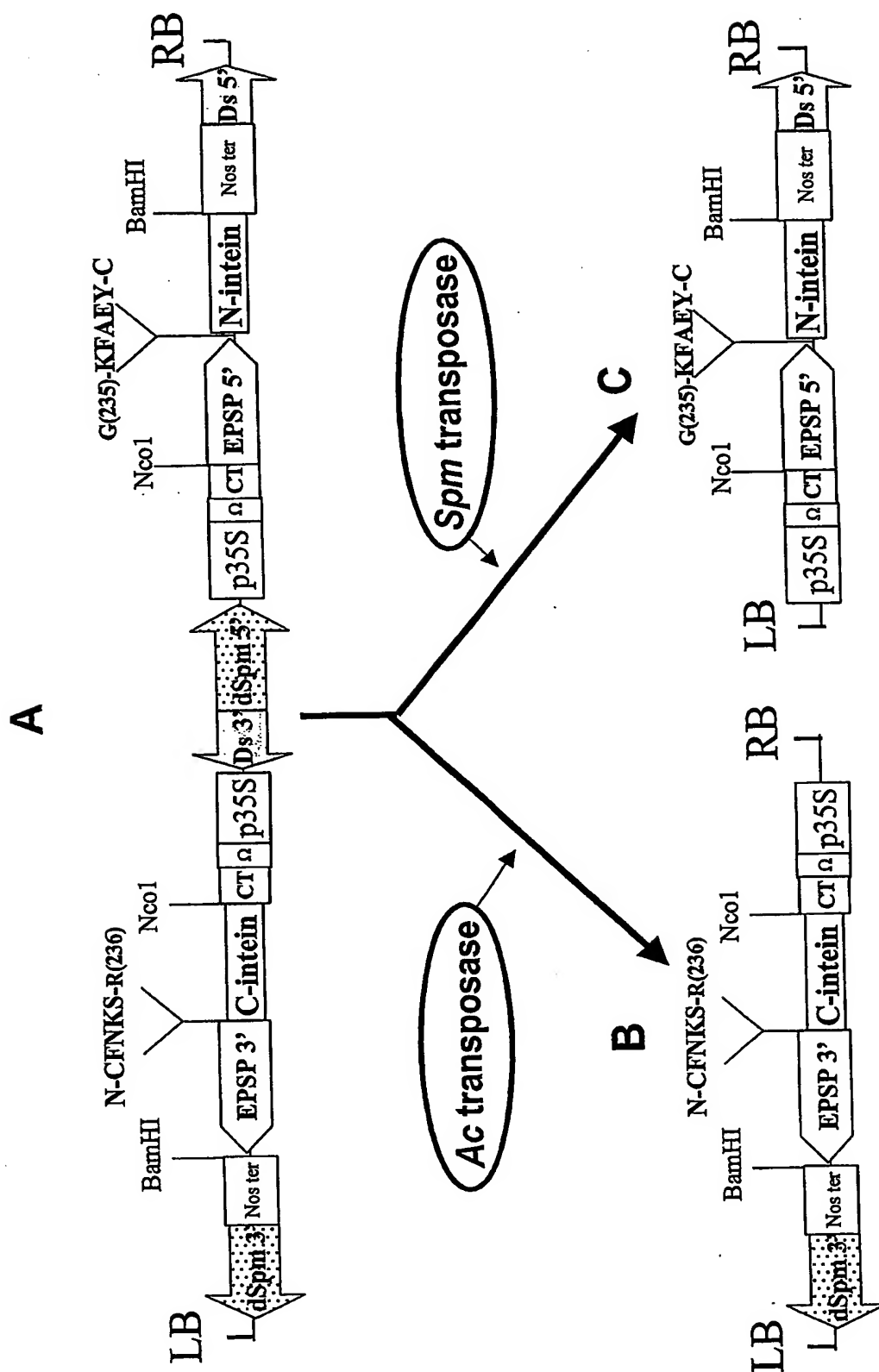


Fig. 6

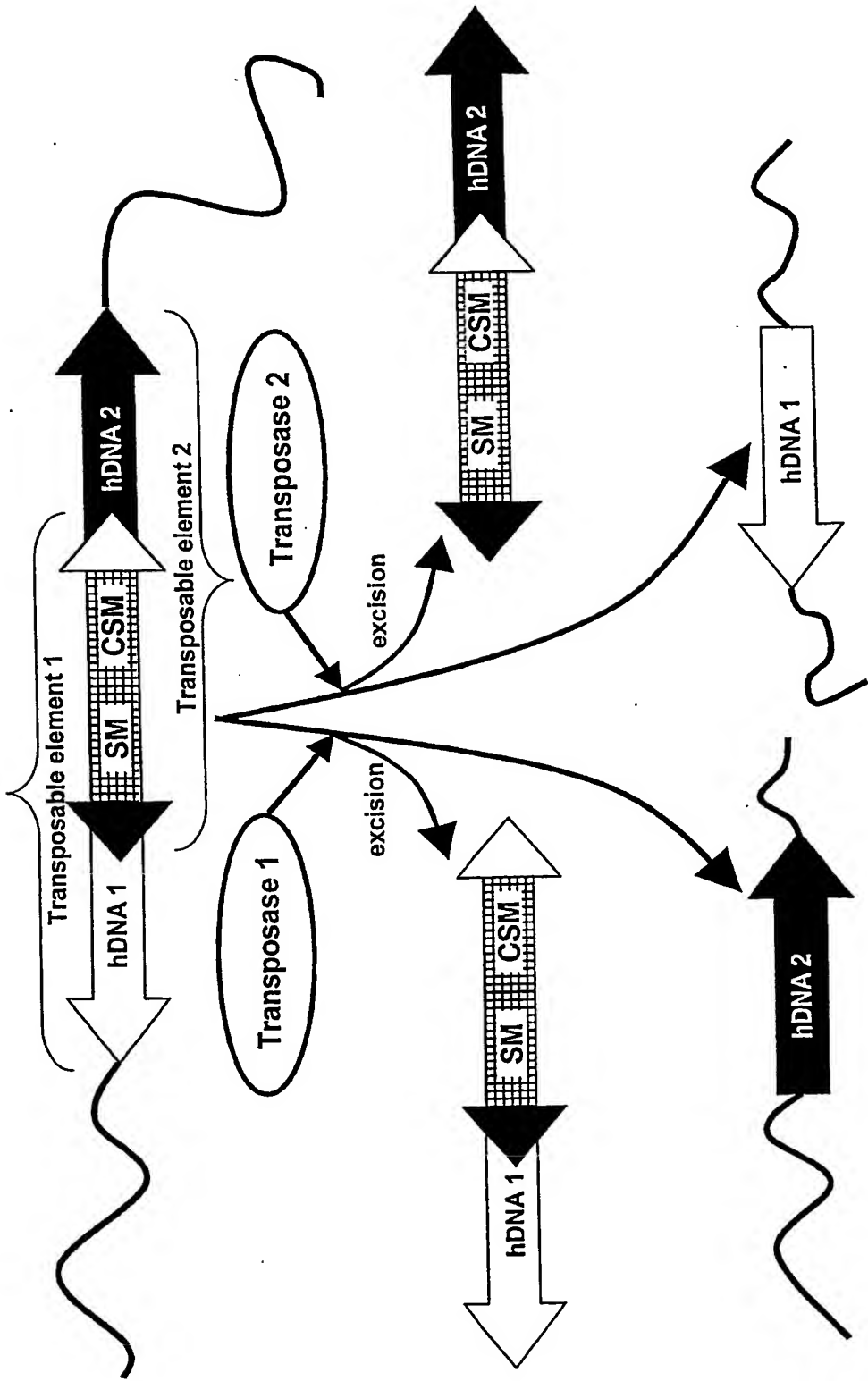


Figure 7

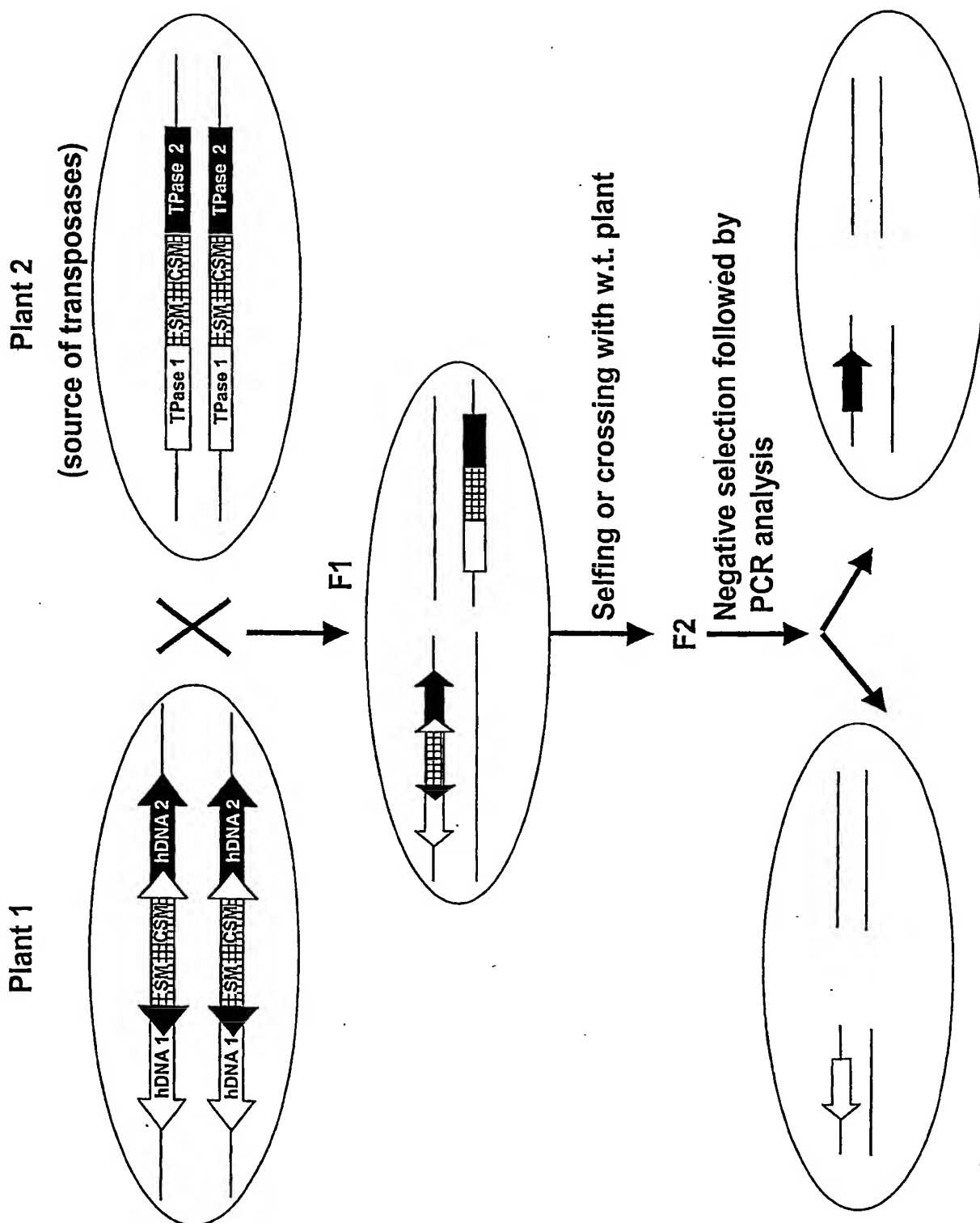


Figure 8

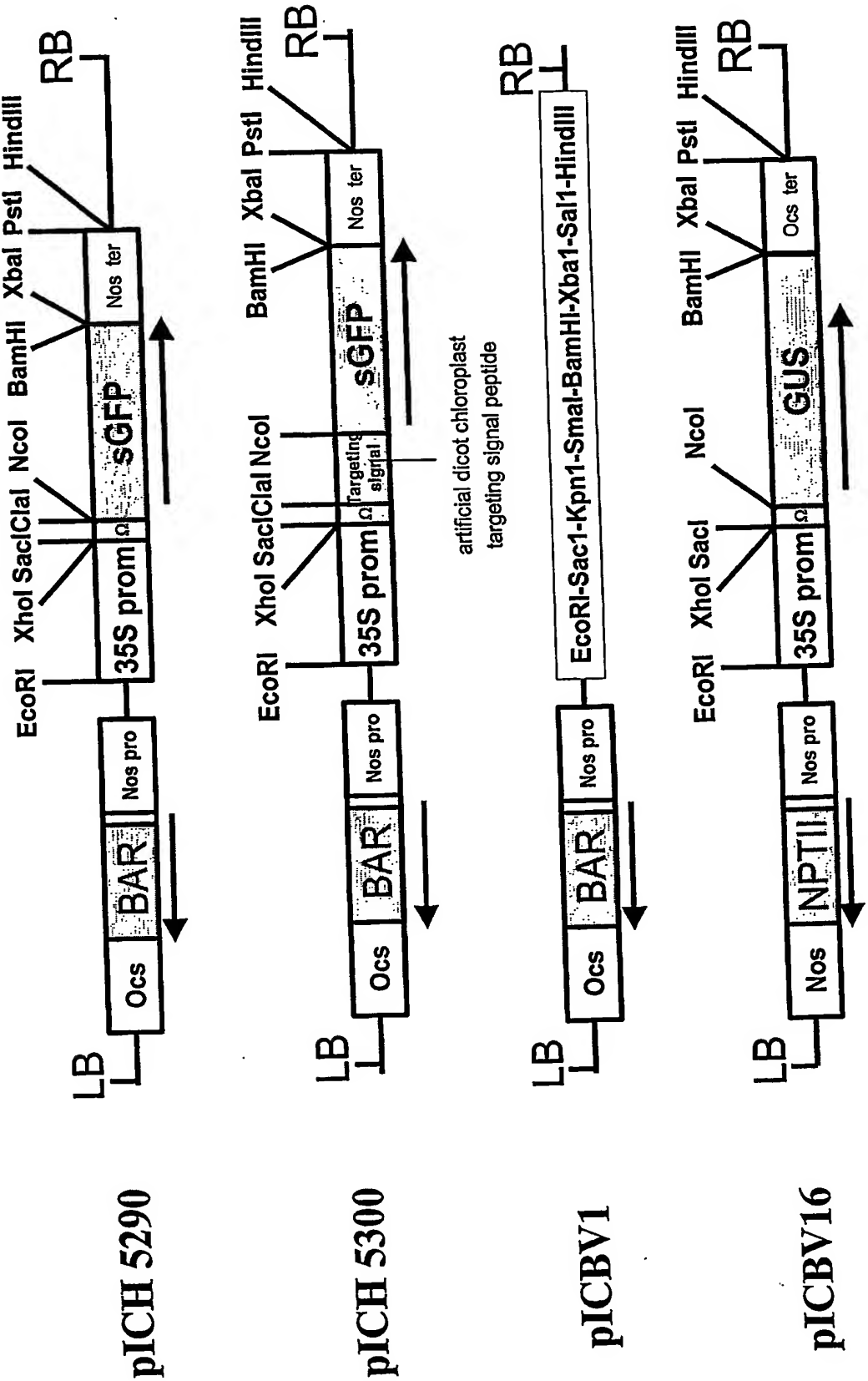


Fig. 9

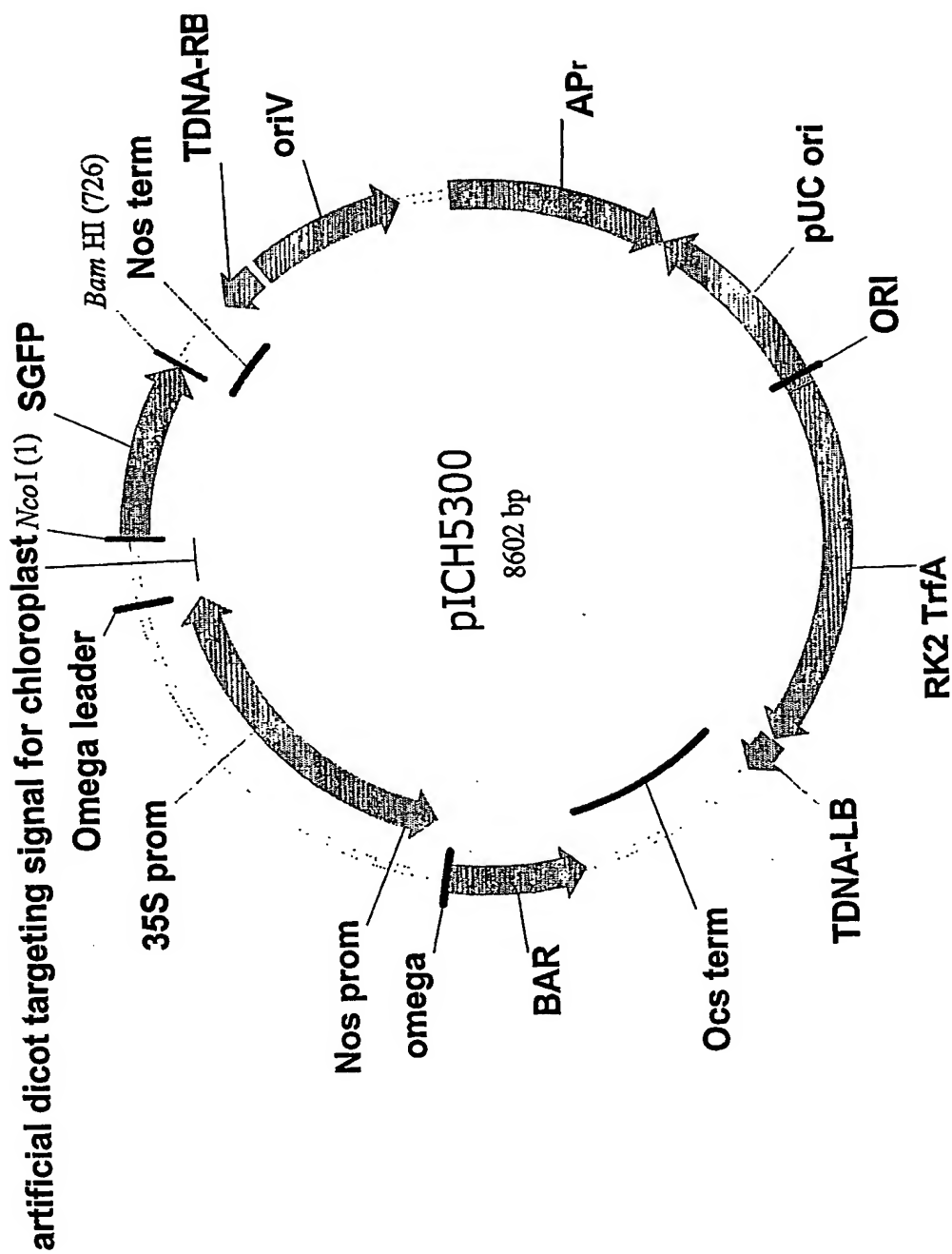


Fig. 10

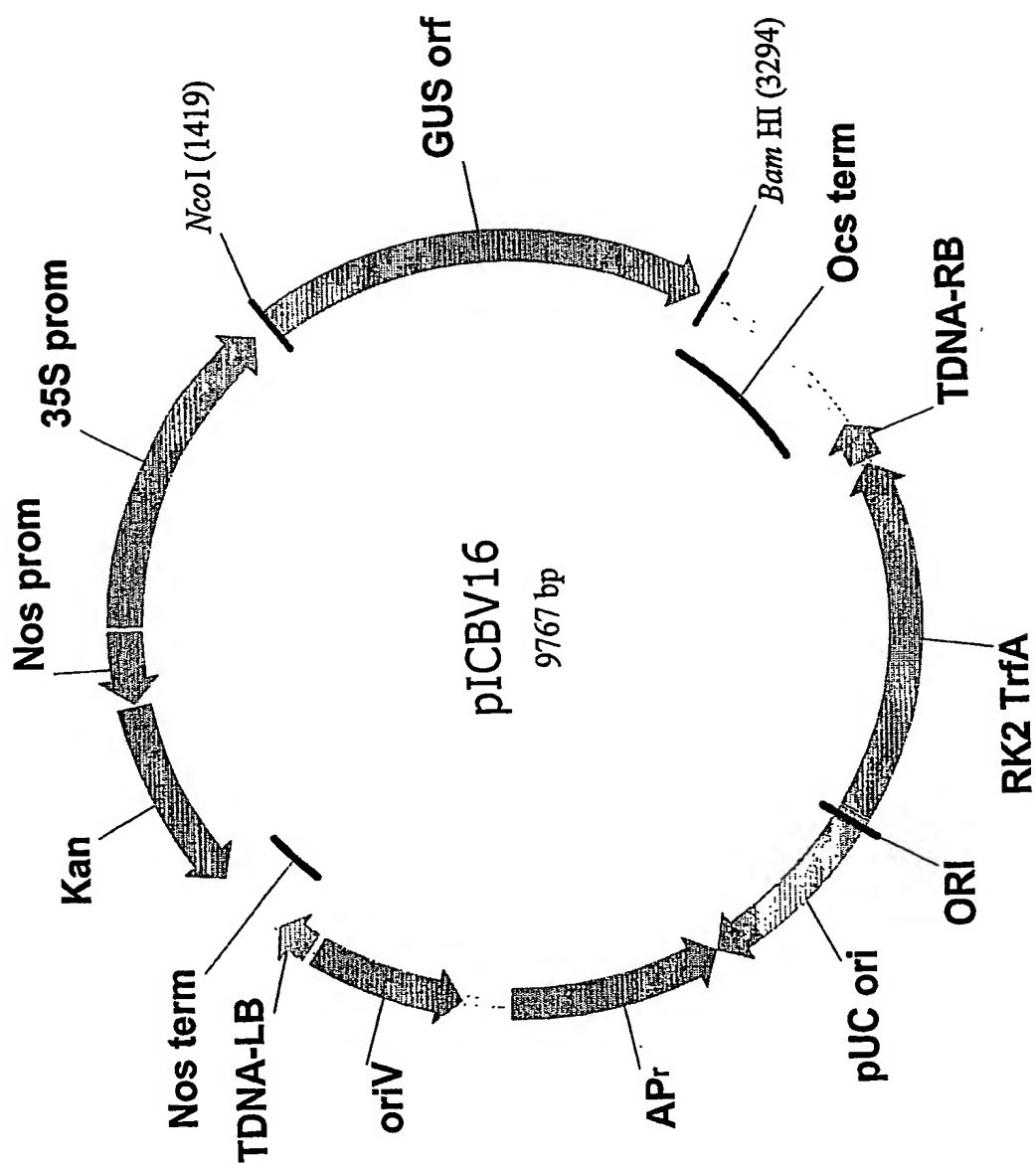


Fig. 11

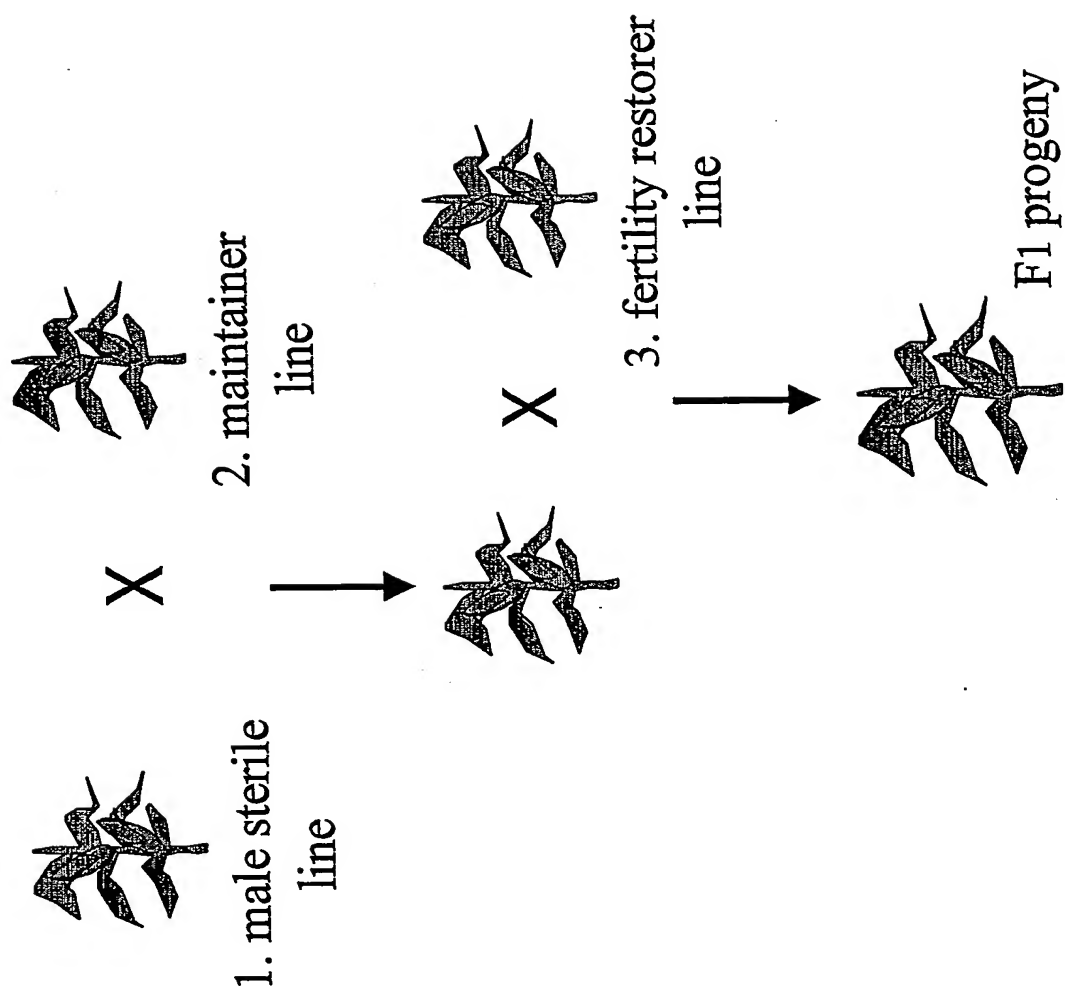


Fig. 12

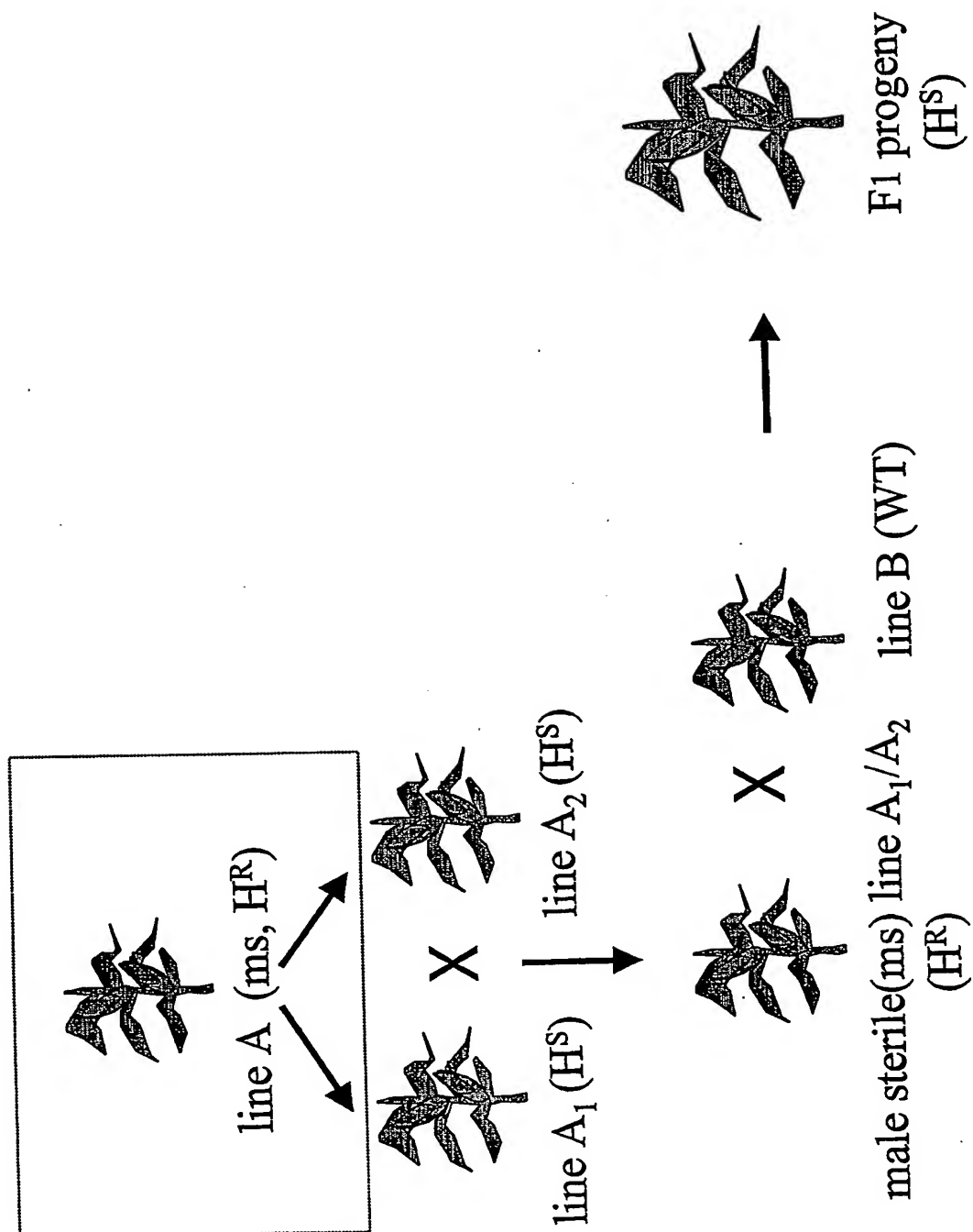


Fig. 13

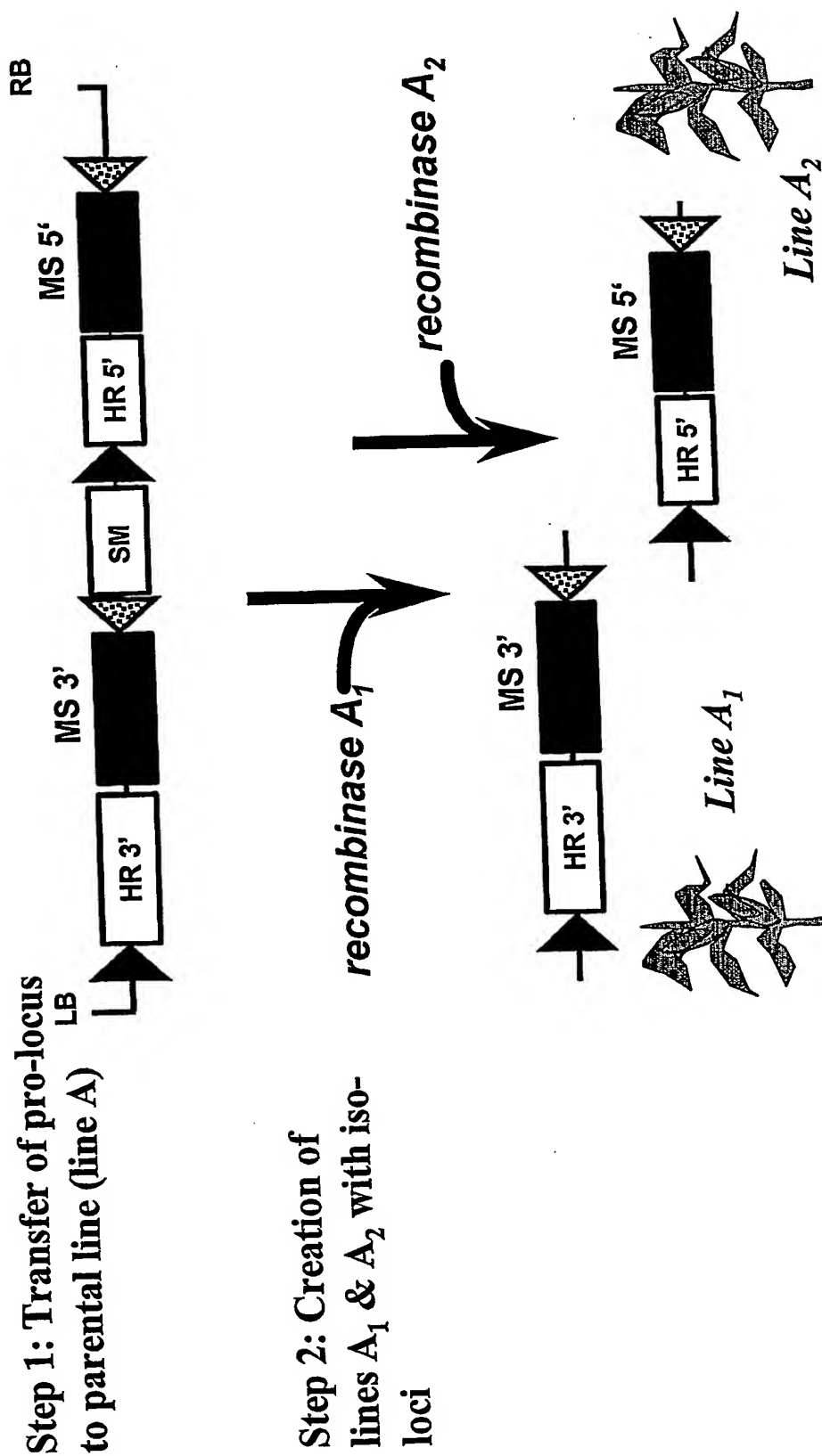


Fig. 14A

**Step 3: Creation of
male sterile parent**

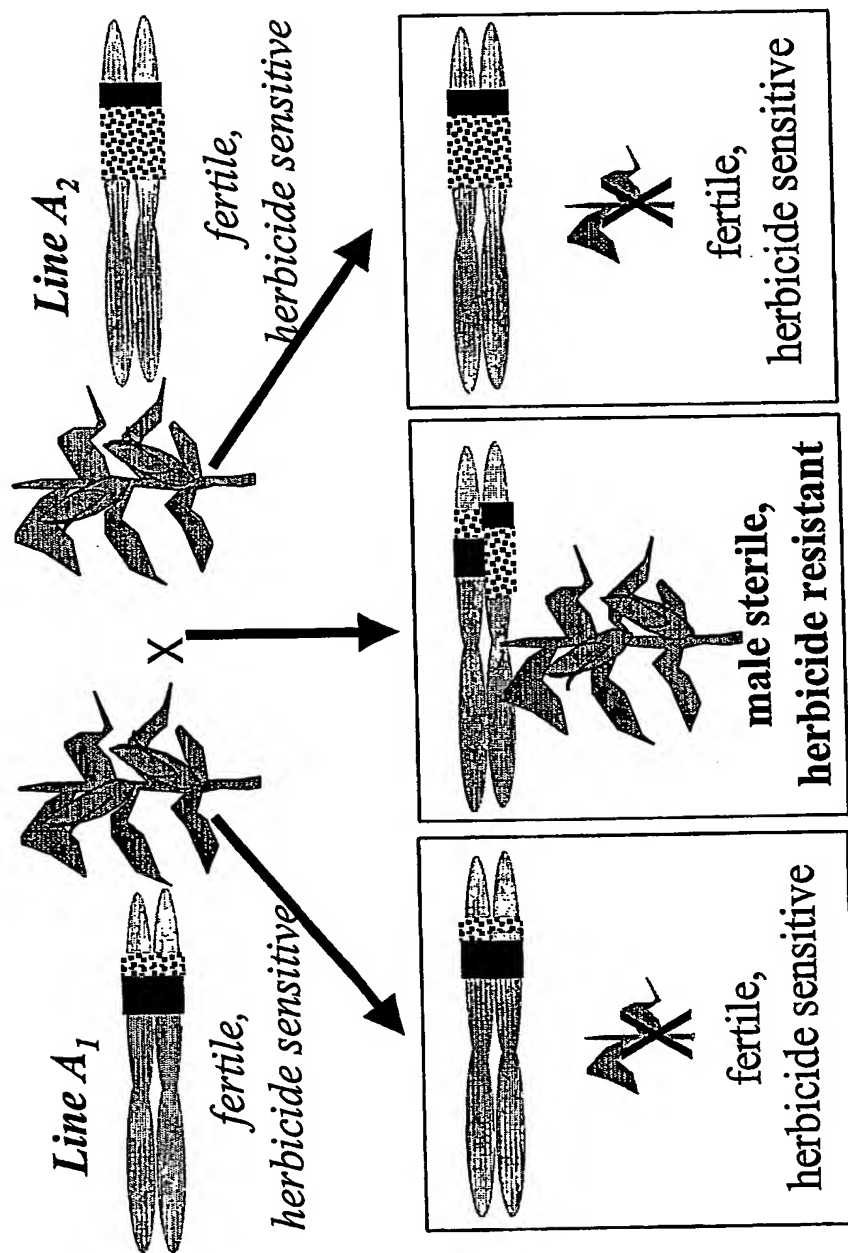


Fig. 14B

Step 4: Production of hybrid seed

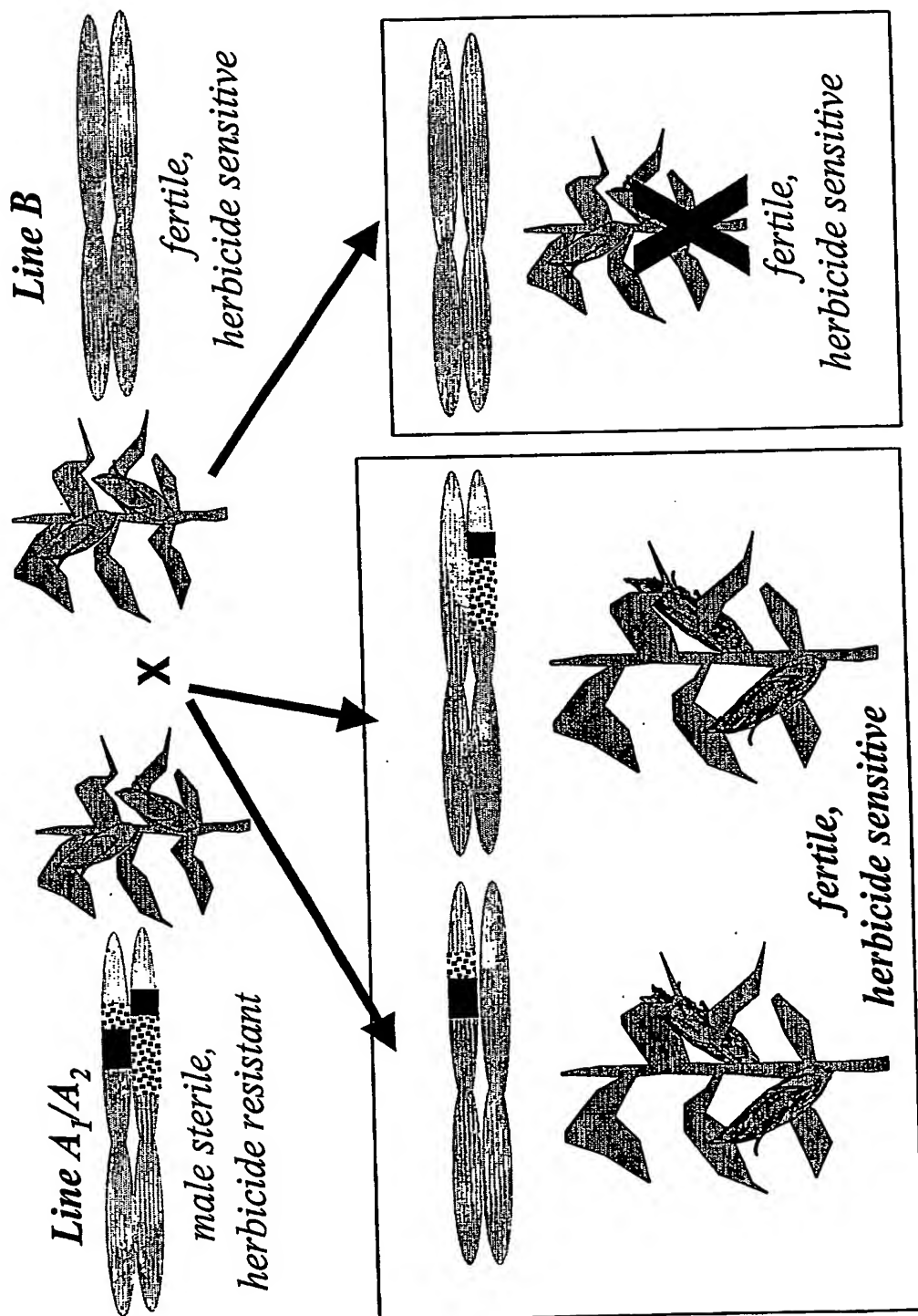


Fig. 14C

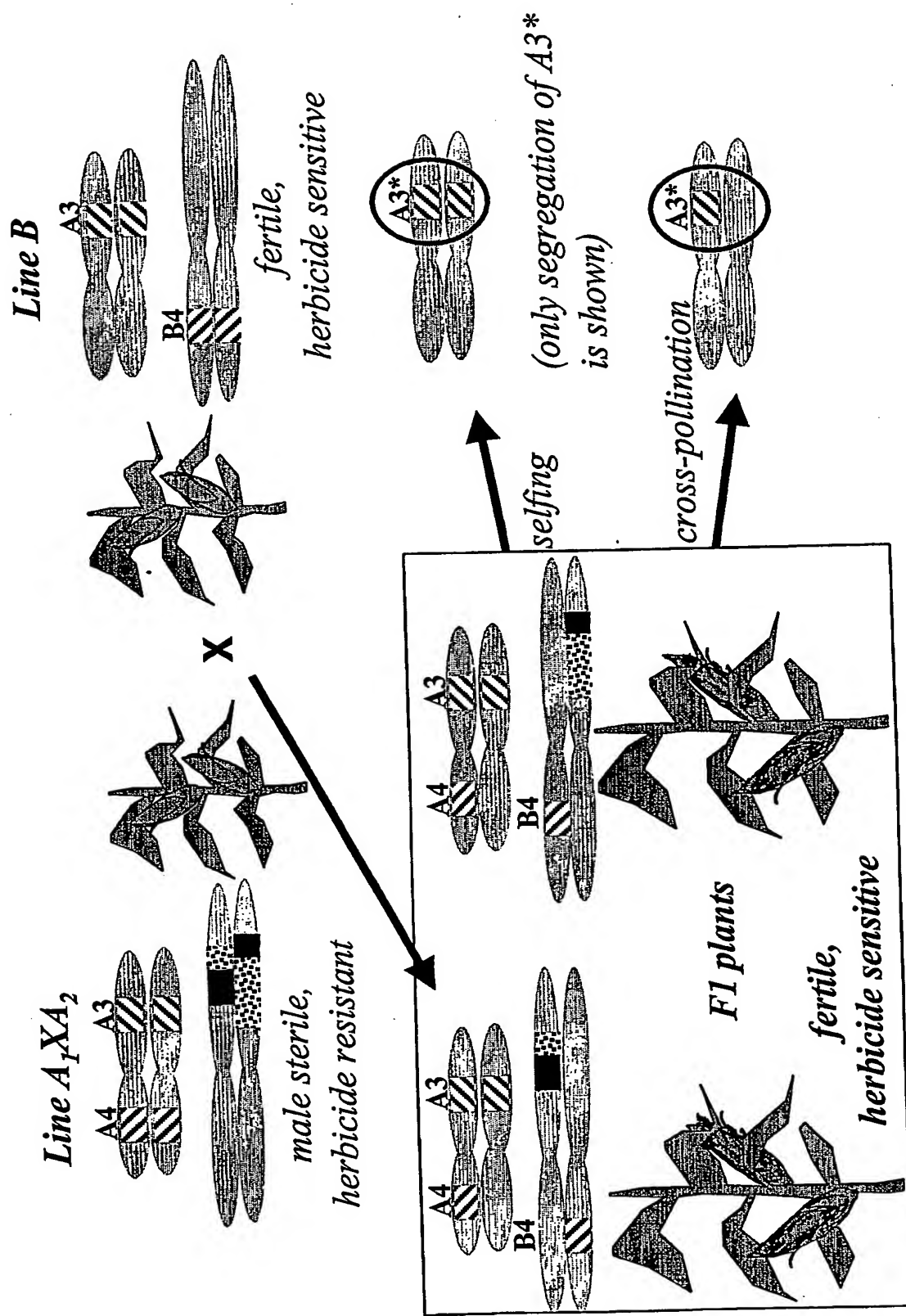


Fig. 14D

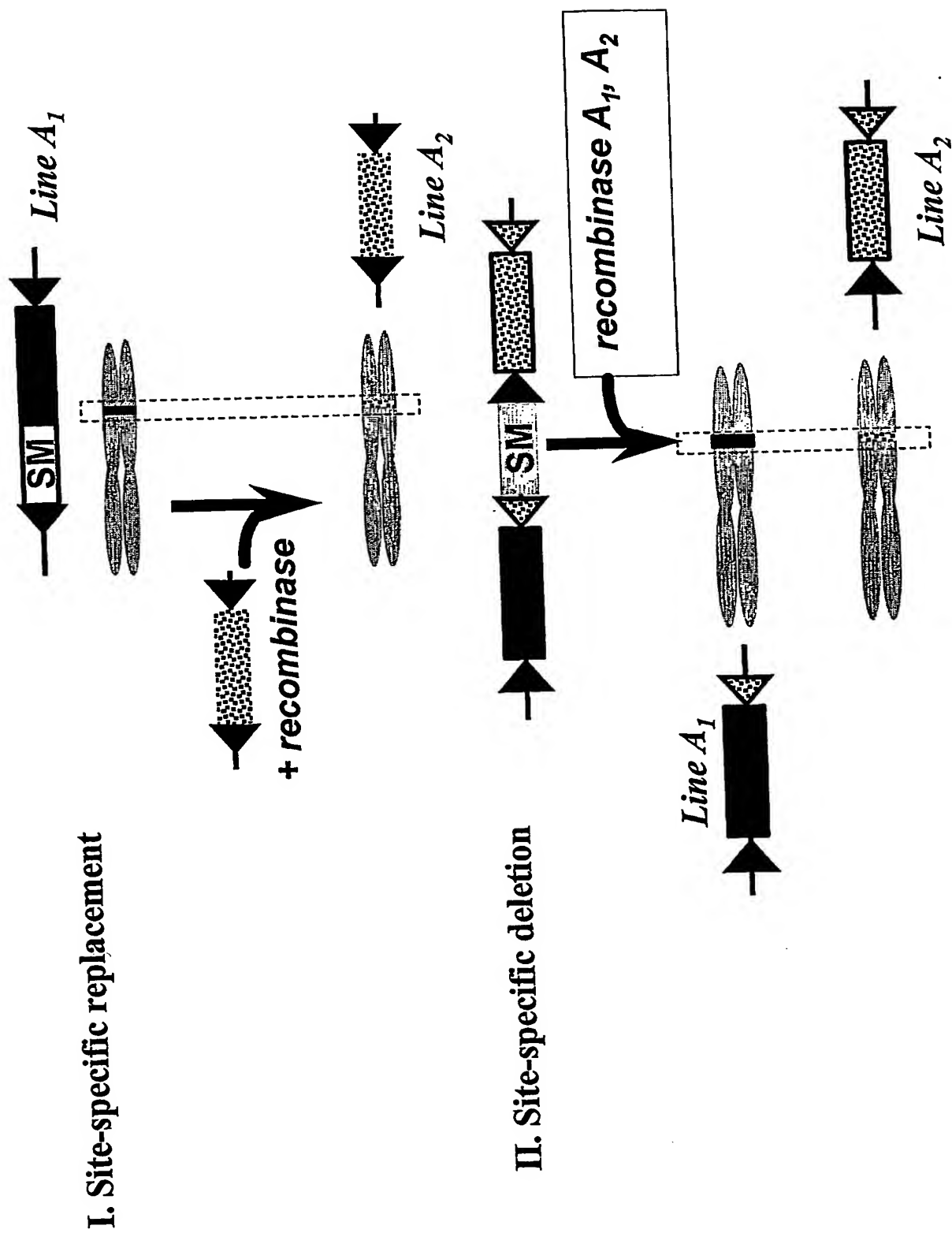


Fig. 15

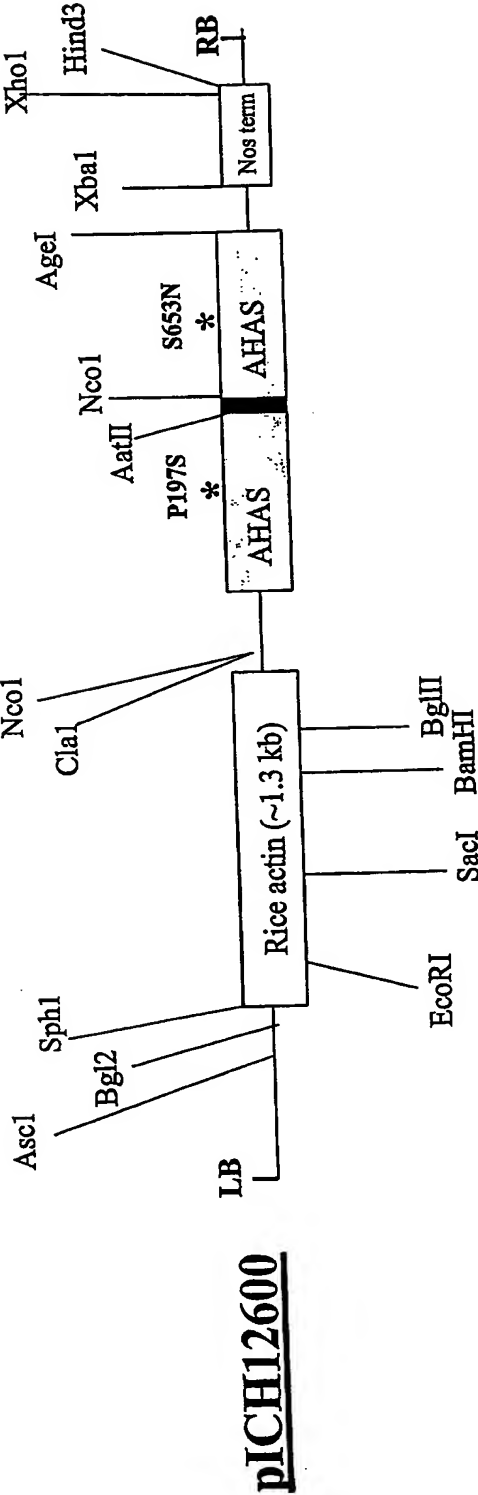
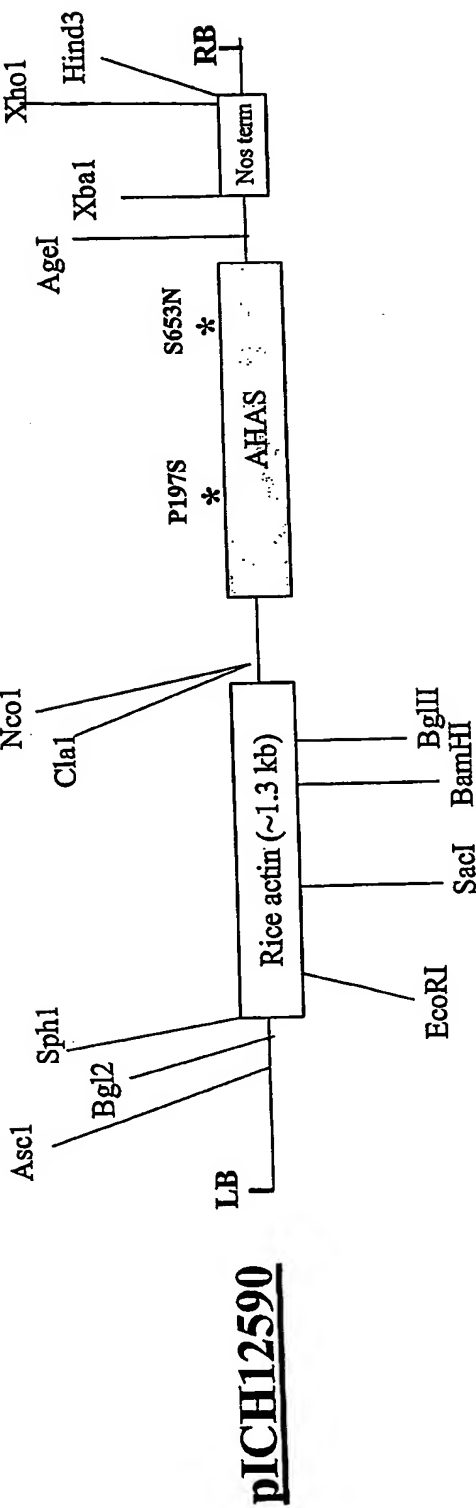
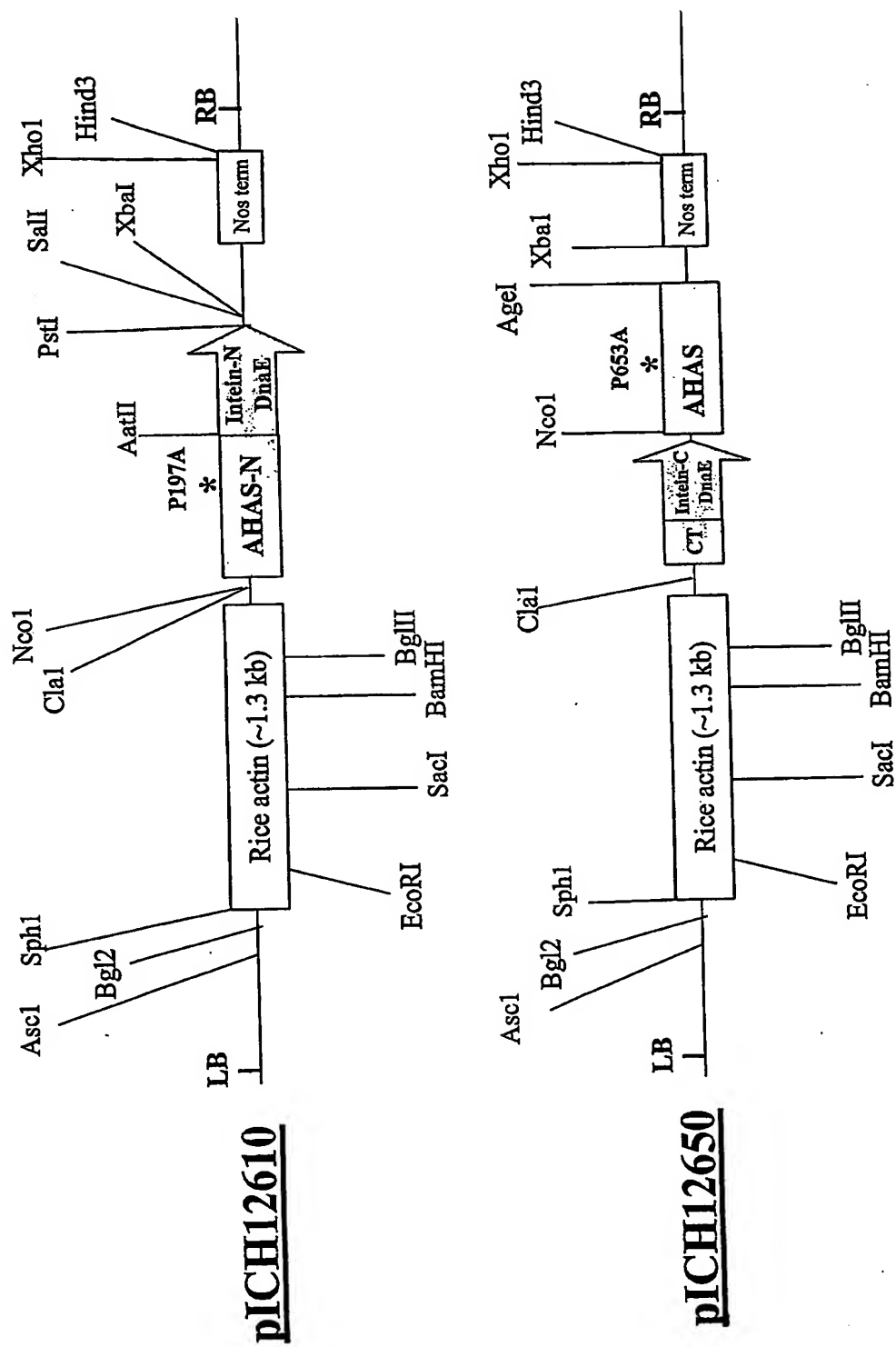


Fig. 16

**Fig. 17**

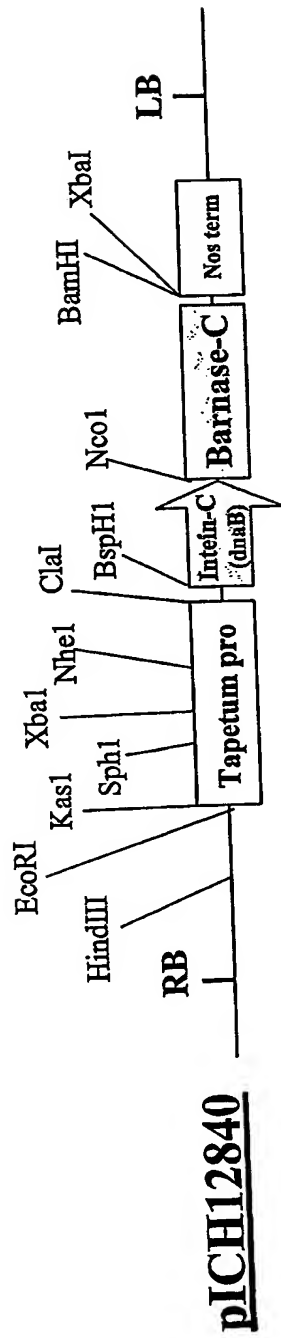
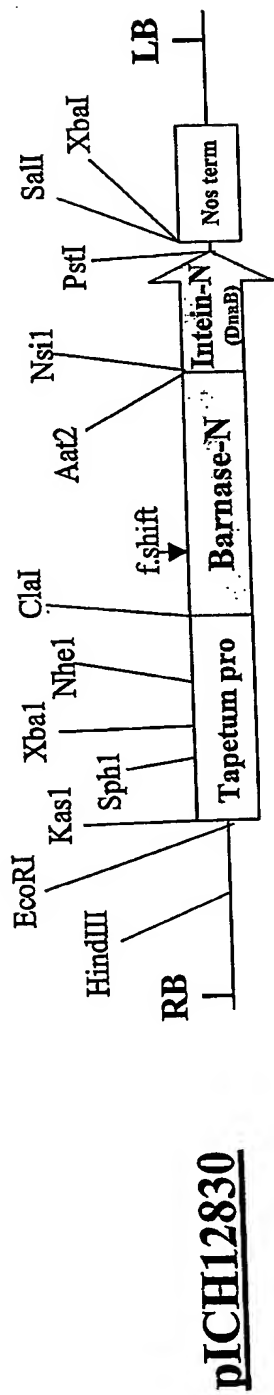


Fig. 18

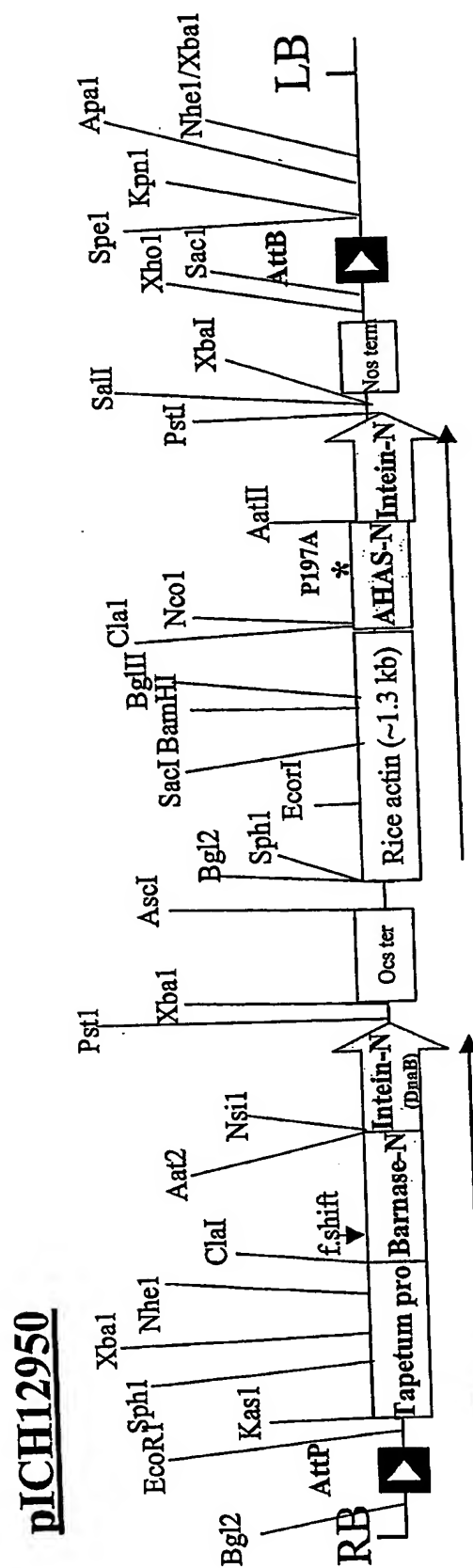
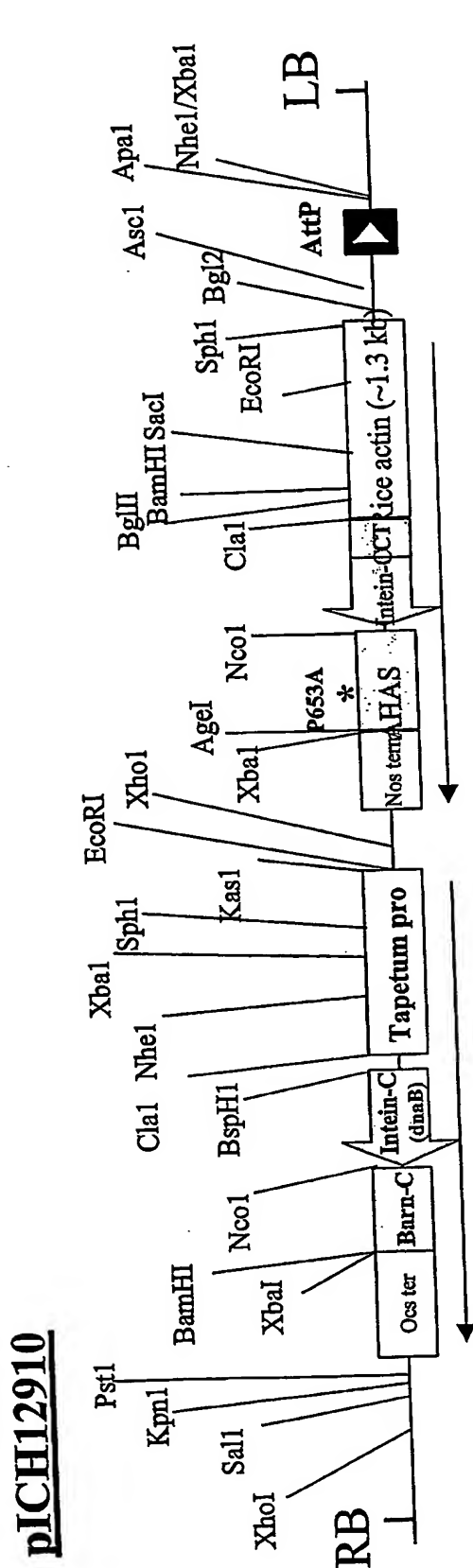


Fig. 19

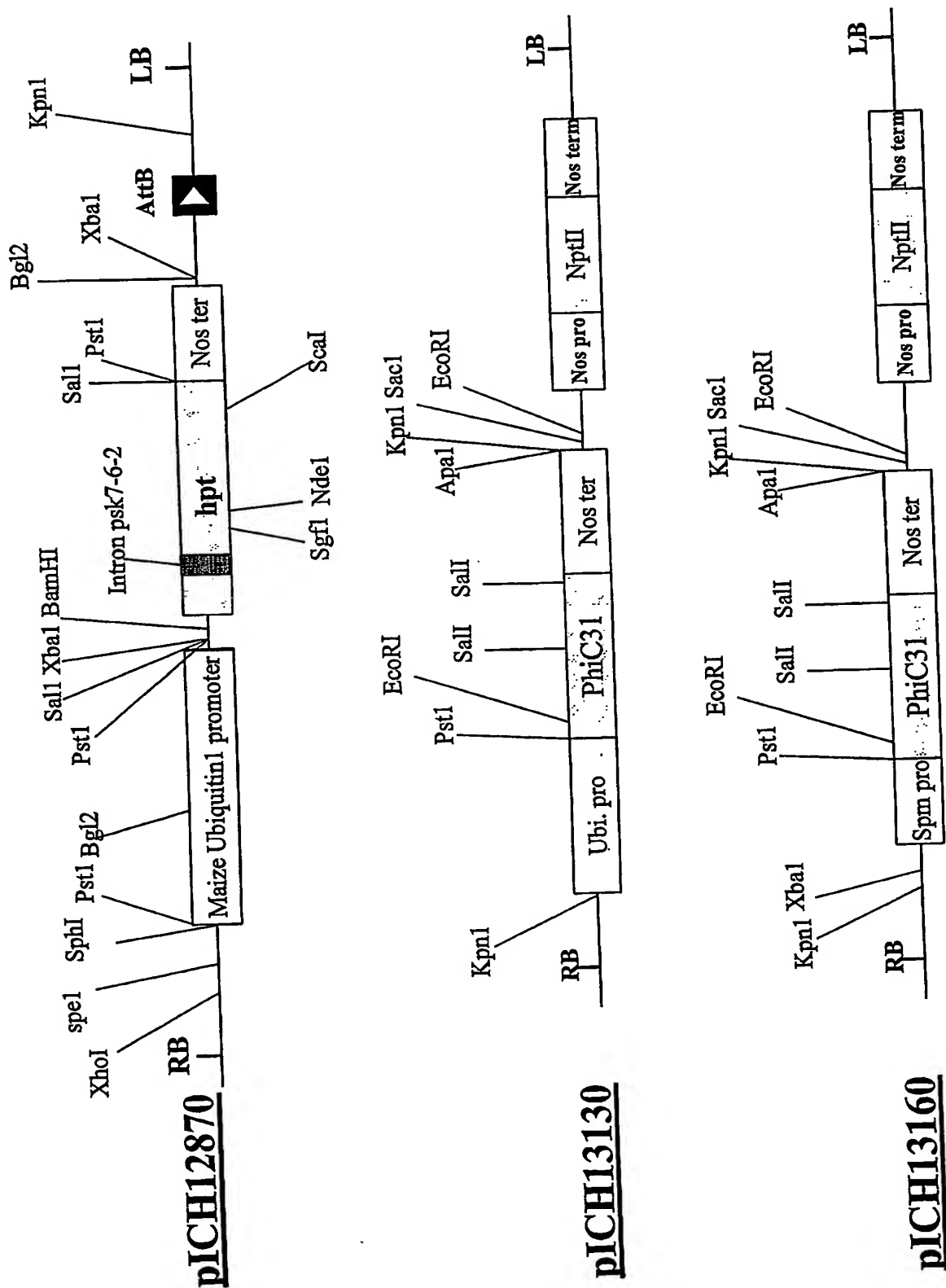


Fig. 20

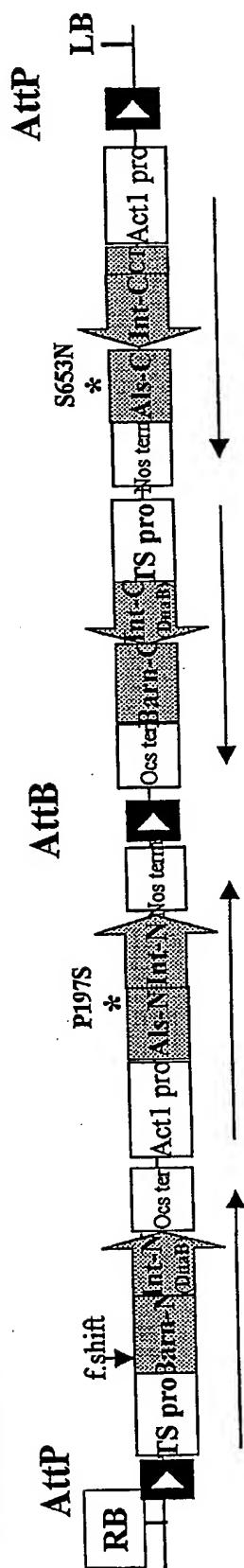
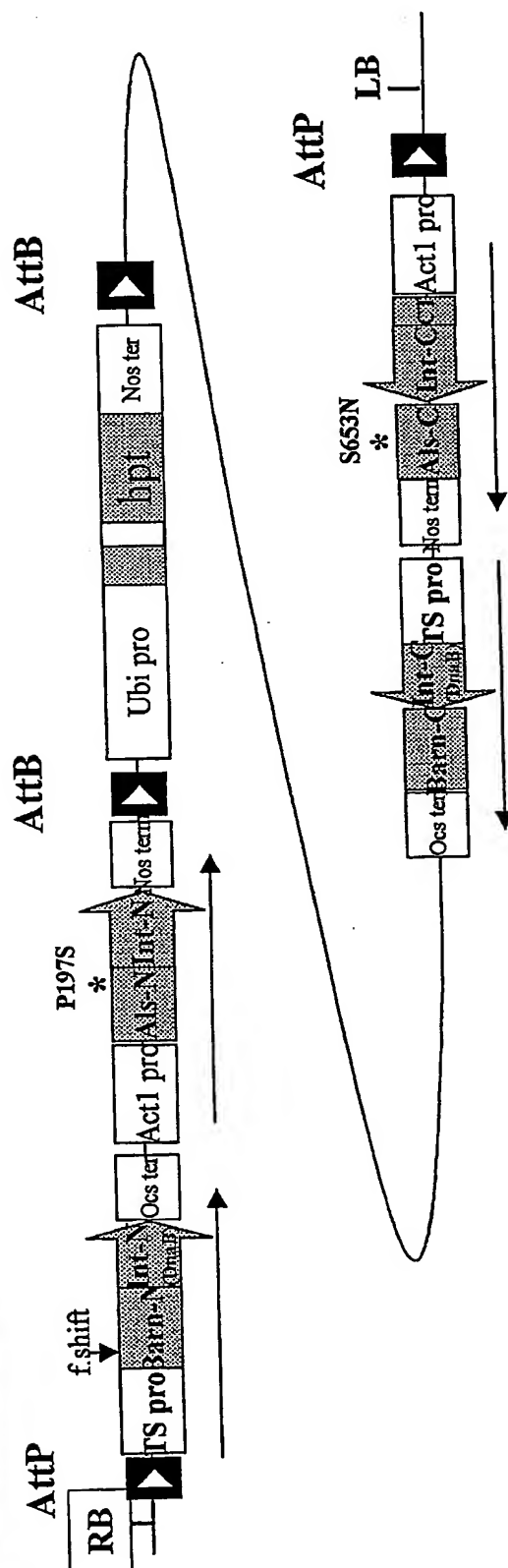
pICH12960pICH12970

Fig. 21

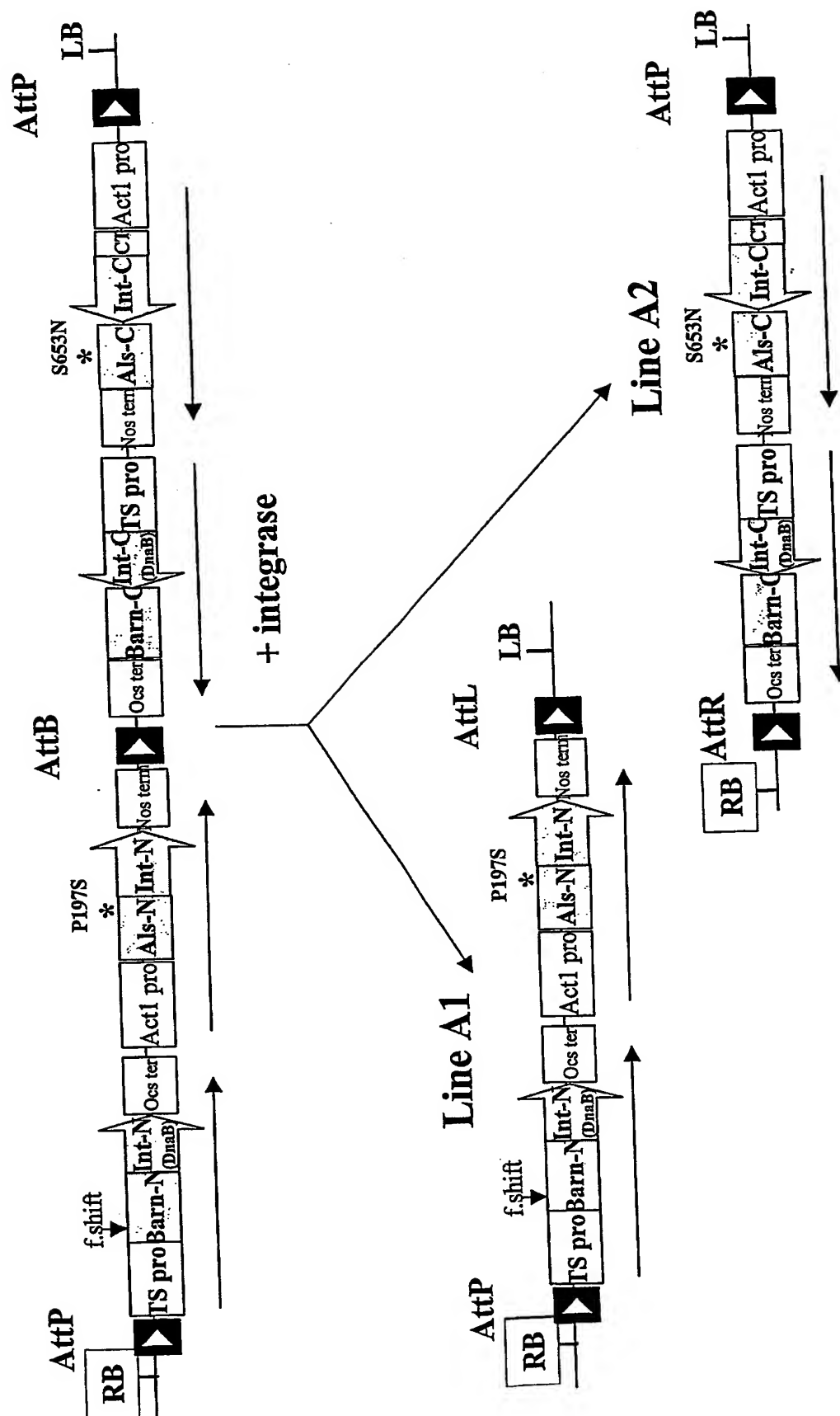


Fig. 22

INTERNATIONAL SEARCH REPORT

Internat Application No

PCT/EP 03/02986

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/82

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, EMBASE, MEDLINE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 00 71701 A (NEW ENGLAND BIOLABS INC ;COMB DONALD G (US); SUN LUO (US); XU MING) 30 November 2000 (2000-11-30) cited in the application the whole document ---	1-57
Y	WO 00 52146 A (BERMUDEZ ERICKA R ;MAXYGEN INC (US); CARR BRIAN (US); LASSNER MICH) 8 September 2000 (2000-09-08) cited in the application the whole document ---	1-57
Y	WO 01 59091 A (METABOLIX INC) 16 August 2001 (2001-08-16) the whole document ---	1-57

-/--

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

g document member of the same patent family

Date of the actual completion of the international search

12 June 2003

Date of mailing of the international search report

23/06/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Chakravarty, A

INTERNATIONAL SEARCH REPORT

Int:
 ial Application No
 PCT/EP 03/02986

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, P	<p>THOMSON J M ET AL: "ARTIFICIAL GENE-CLUSTERS ENGINEERED INTO PLANTS USING A VECTOR SYSTEM BASED ON INTRON- AND INTEIN-ENCODED ENDONUCLEASES" IN VITRO CELLULAR & DEVELOPMENT BIOLOGY. PLANT, GAITHERSBURG, MD, US, vol. 38, no. 6, November 2002 (2002-11), pages 537-542, XP001149513 ISSN: 1054-5476 the whole document</p>	1-57

INTERNATIONAL SEARCH REPORT

Inventor's Application No

PCT/EP 03/02986

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0071701	A	30-11-2000	AU 5039500 A	12-12-2000
			CA 2374497 A1	30-11-2000
			CN 1350582 T	22-05-2002
			EP 1183346 A1	06-03-2002
			JP 2003505012 T	12-02-2003
			WO 0071701 A1	30-11-2000
WO 0052146	A	08-09-2000	AU 3391900 A	21-09-2000
			AU 3719500 A	21-09-2000
			CA 2362737 A1	08-09-2000
			CA 2364997 A1	08-09-2000
			EP 1165757 A1	02-01-2002
			EP 1165775 A2	02-01-2002
			WO 0052146 A2	08-09-2000
			WO 0052155 A2	08-09-2000
			US 6365377 B1	02-04-2002
			US 6531316 B1	11-03-2003
			US 6406910 B1	18-06-2002
			US 6358740 B1	19-03-2002
			US 6413745 B1	02-07-2002
WO 0159091	A	16-08-2001	AU 3683901 A	20-08-2001
			EP 1255846 A2	13-11-2002
			WO 0159091 A2	16-08-2001
			US 2002129400 A1	12-09-2002

which was changed from Cys to Ala. PIC5'epsp-intM was made following the same strategy as for PIC5'epsp-int except that primer intepr7 (caa cag tat cac tct cca ggt ttt gcg gaa tat gcc ctc agt ttt ggc ac) was used instead of primer intepr5.

PICint-epsp3'M is similar to construct PICint-epsp3' but differ at the junction C intein-EPSP by the addition of 3 C extein amino acids instead of five, the first one mutated from Cys to Ala and the two other native, and by the last C intein aminoacid which was changed from Asn to Ala. PICint-epsp3'M was made following the same strategy as for PICint-epsp3' except that primer intepr8 (atc gcc ctc gac cag ata gcg gtt aaa agc agc ggc ggc gat cgc ccc att g) was used instead of primer intepr6.

The three mutated aminoacids completely prevent intein mediated transplicing but do not prevent association of the N and C intein fragments ((Chen et al., 2001, *Gene*, 263, 39-48). pIC5'epsp-intM and pICint-epsp3'M were transformed in agrobacterium strain GV3101 by electroporation. Both constructs were transformed in *Arabidopsis thaliana* (Col-0) and tobacco as described above. Primary transformants were all sensitive to glyphosate, but hybrid F1 plants containing both constructs, either in tobacco or *Arabidopsis*, exhibited glyphosate resistance.

EXAMPLE 4

Splitting the Arabidopsis AHAS gene

The acetolactate synthase (AHAS) gene from *Arabidopsis* (Genbank accession AY042819) was amplified from *Arabidopsis* genomic DNA using primers *Als1* (5' taaacatgg cgccggcaac aacaac 3') and *Als2* (5' gactctagac cggtttcac tctcagtatt taatc cgcc atctcc 3') and cloned as an Nco1-Xba1 fragment in Icon Genetics binary vector pICBV24 (Kan^R, selection in *E.coli* and *Agrobacterium*). Ser653 was mutated to Asn by PCR using primers *Alsm5* (5' caggacaagt ctctcgtcgt atg 3'), *Als4* (5' gaaagtgcc ccatcgga tcatcg 3'), *Als3* (5' cgatgatccc gaatggtggc ac 3') and *Als2*. The amplified mutated fragment was cloned as an Nhe1 - Age1 fragment. A second aminoacid, Pro197 was mutated to Ser by PCR using primers *Als1*, *Alsm5*, *Alsm6* (5' acgacgagag actgtcctg tg 3') and *Alspr1*. The amplified mutated fragment was subcloned as a Sapl-MluI fragment.

The rice actin1 promoter was amplified by PCR from rice genomic DNA using primers *Actpr1* (5' atgggcgccg cagatctgca tgccggtcga ggcatcat atgcttgag 3') and *Actpr2* (5' cgccatgggtt

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.